THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR .... Betsy Terry ....

TITLE OF THESIS .... An Evaluation of the Generalization of Motor ....

.... Skills in the PREP Program ....

DEGREE FOR WHICH THESIS WAS PRESENTED .... Masters of Science ....

YEAR THIS DEGREE GRANTED .... 1981 ....

THE UNIVERSITY OF ALBERTA

AN EVALUATION OF THE GENERALIZATION OF MOTOR SKILLS IN THE PREP PROGRAM

BY

© Betsy Terry

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTERS OF SCIENCE

DEPARTMENT OF PHYSICAL EDUCATION

EDMONTON, ALBERTA

SPRING, 1981

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled "An Evaluation of the Generalization of Motor Skills in the PREP Program"
submitted by Betsy Terry
in partial fulfillment of the requirements for the degree of Masters of Science.

DEDICATION

"If I have been able
to see farther,
it was because
I stood on the shoulders of giants."
Newton

To all those giants-
far and near,
great and small.

ABSTRACT

A category observation instrument was designed to capture the range of behaviors related to generalized skill performance. A criterion observer and two other trained observers utilized this instrument to record the free play behavior of eight moderately mentally retarded children. The instrument was used to transcribe the information provided on videotape recordings into descriptive free play data for the eight subjects.

During the two months of filming, the subjects received instruction in six different play skills. A multiple baseline design across behaviors in combination with an ABA design was employed.

One dependent variable was carefully chosen from nine possibilities, to measure the generalization of active free play behavior in PREP. Criteria were established to facilitate the evaluation of behavior change. Examination of performance graphs for individual subjects, with respect to this criteria revealed that some generalized behavior was observed for all subjects. The extent of change was variable across subjects.

Analysis of the generalization of specific skills, based on performances of seven skills prescribed to three or more children, indicated that generalization of some skills may be more pronounced than others. It was concluded that generalization may be somewhat skill specific.

ACKNOWLEDGEMENTS

I would like to thank the following folks who have helped to make this thesis become a reality . . .

Donna, Anne, Maureen, Heather, and Mara for your unending patience with runny noses, 3 second "bleeps", and all my requests,

My family and friends, some who have given constant support by accepting my absence and others who have freely given the friendship and encouragement needed to see the light at the end of the way,

with special thanks to Carol, Maurie, and Pat,

The teachers and children from Winnifred Stewart School who have left me with fond memories of my days spent in PREP,

Dr. Jane Watkinson for your advice and reassurance from the first to the last of my days here at the University of Alberta,

Dr. Ted Wall and Dr. Gerry Kysela for sharing your wealth of knowledge with me, assisting in the writing of this thesis, and serving on my committee.

Thank You

TABLE OF CONTENTS

LIST OF TABLES

LIST OF DIAGRAMS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Play can be thought of as a fundamental component of life, a phenomenon basic to learning. Although a variety of play theories exist, it is generally accepted that play is a vehicle for learning (Weininger, 1964), especially for children. The importance of play to healthy growth and development physically, socially, emotionally, and cognitively is supported by the following well respected researchers in the study of play behavior:

> "Play behavior is crucial, varied, and individual..."
> (Ellis, 1976, p.137)

> "It seems obvious from the observations... that play is not the simple time filler it was once considered to be. Indeed, it is perhaps the most important way in which the physical, emotional, and intellectual development of a child is advanced." (Weininger, 1972, p.64)

> "Living in a state of play means living more humanly."
> (Levy, 1978, p.1)

In light of these comments, recent increased interest in the play phenomenon is not surprising. Concise and systematic observation of children's play patterns provides evidence concerning the efficient or inefficient use of free play time. Formal observation of mentally retarded youngsters (Horne and Philleo, 1942; Keeran, Grove, and Zachofsky, 1969; Linford, 1971; Noble, 1975; Peterson and Haralick, 1977; Strain, 1975; Tilton, 1964; Wall, 1974; Watkinson, 1977) demonstrates that play patterns of mentally retarded children are inferior to those of their nonretarded peers. Not only do mentally retarded children often have less free play time available to them, but when they do have the time they do not use it constructively. A deficit is seen in the amount and sophisti-

cation of play skills initiated. A large portion of their free play time is spent in non-play.

Certainly play is as important to the development of the mentally retarded child as any child. For this reason, it becomes necessary and vital to develop appropriate teaching materials and strategies to meet the special needs of young mentally retarded children.

Several different approaches can be taken in an attempt to improve the play skills of the mentally retarded. One approach includes simply increasing play time and equipment opportunities. Some programs employ strategies for teaching and encouraging modeling and imitation, as a possible means of upgrading free play performance. Other programs utilize direct teaching models developed specifically for the mentally retarded. The PREP Program (Watkinson and Wall, 1980) at the University of Alberta, Edmonton is based on the last approach.

The PREP Play Program began in 1974. Since then, many revisions have been made but the goal of the program remains unchanged- to teach young moderately mentally retarded children to play. As outlined in the PREP manual (Watkinson and Wall, 1980) the three features of the program are individualized instruction on specifically prescribed play skills, small group activities, and opportunity for free play in a stimulating play environment.

The backbone of the PREP Program is a collection of task-analyzed culturally normative play skills. Following a highly specific instructional model (Watkinson and Wall, 1980), individual instruction in these skills is interjected into free play time. It is the teacher's goal to increase the initiation rate of proficient play behaviors. PREP strategies have been successful in improving the play skill level of moderately

mentally retarded children (Noble, 1975; Shatz, 1979, Watkinson, 1977). In order to further justify the teaching strategies used in the PREP Program it becomes necessary to formally look at a relationship between the skills being taught to a child in individual instruction and the child's choice to initiate those specific skills in free play.

Statement of the Problem

It was the purpose of this study to evaluate the generalization into free play, of skills receiving direct instruction in the PREP Program. VTR equipment and an appropriate category observation instrument (Terry, 1980) were used to accomplish this task.

More specifically, the following questions were asked:

1. Using a category observation instrument designed to describe a range of behaviors that could be considered as generalized training effects, which category or group of categories provide(s) a realistic and conservative measurement of generalization?
2. To what extent do behaviors which have been taught and programmed to generalize, actually generalize into free play in the PREP Program?

Limitations

1. Children included in the study were selected on the basis of their participation in the PREP Program and therefore the extent to which they are representative of moderately mentally retarded children is unknown.

2. Due to the small sample size of eight, the generalizability of results to other moderately mentally retarded children outside of the PREP Program is limited.

3. This study may have been limited to a minor extent in that three cameras were placed in the PREP room. However, considerable effort was made to ensure that the filming was unobtrusive.

4. As teachers were trained to use the curriculum materials specific to this study, it must be assumed that all teachers used the prompting system along with reinforcement and feedback as described to move the children through the task sequences.

5. The criterion observer and two secondary observers were also involved in the teaching portion of this study. However, video tape recordings were coded in a random order to minimize observer bias.

Delimitations

1. This study was delimited to a sample of eight students who were participants in the PREP Program.

2. The treatment phase of the study included fourteen instructional sessions during the period from May 7 to June 23, 1980.

3. The investigation included only those skills prescribed to the eight subjects on the basis of three hours of free play observation and PREP guidelines.

Definitions

The following terms have meanings specific to the purposes of this study:

1. Free play time. This term refers to the time available to the child in the PREP Program when he is free of teacher interaction.

2. Individualized Instruction. This term describes a brief instructional episode in which a teacher interacts with one child, employing systematic teaching techniques, designed to assist the child to perform independently as early as possible.

3. Maintenance. This term refers to the systematic attention given to a target skill in an attempt to increase the initiation of the skill in free play.

4. Task Analysis. Task analysis refers to the breaking down of a complex skill into its component behaviors and arranging them in an order of difficulty so that attainment in a sequential order leads to performance of a target skill.

CHAPTER II

REVIEW OF LITERATURE

Generalization

Instructional programs set out to change behaviors, sometimes increasing and other times decreasing their frequency, depending on the nature of the behaviors under study. It seems logical to assume that a successful program results not only in the mastery of certain skills but also, and perhaps even more importantly, in the generalization of these newly acquired skills.

Definitions of generalization in the literature vary but each one suggests a transfer of learning from the restraints of the experimental setting to typical everyday situations (Kazdin and Bootzin, 1972; Stokes and Baer, 1977; Wehman, Abramson, and Norman, 1977). The concept of generalization is more clearly understood by dividing it into two classifications: response generalization and stimulus generalization (Kazdin and Bootzin, 1972). Response generalization describes an improved performance of desirable behaviors other than those initially identified as target behaviors. Stimulus generalization suggests that new behaviors persist under conditions other than those in which they were originally learned. The precise meaning of stimulus generalization depends upon the treatment conditions being considered. For example, it can be a transfer of performance from one setting to another, from the control of one teacher to another, or from the control of extrinsic reinforcement to naturally occurring reinforcers.

Quite clearly then, stimulus generalization should be a desired

outcome of any instructional or behavioral program. After all, how effective is a program that produces change only during intervention? It is fair to say that stimulus generalization is indeed a desired outcome of nearly every behavioral researcher. However, more than half of the studies reviewed by Stokes and Baer (1977) chose to do nothing more than "hope" for a generalized effect. This is rather surprising in light of the fact that Baer, Risley, and Wolfe (1968) indicated several years prior to many of these investigations that generalization must be programmed rather than expected.

Studies over the past ten years invariably concluded that "unprogrammed" generalization just did not occur (Jackson and Wallace, 1974; Rincouver and Koegel, 1975; Stokes, Baer, and Jackson, 1974; Walker and Buckley, 1972; Walker, Hops, and Jackson, 1975). Failure to bring about change in fundamental behaviors such as social interaction skills, language development, and self help skills gradually prompted serious investigation into possible techniques for the programming of generalization.

Techniques for Increasing Generalization

As behaviorists began to regard generalization as a behavior in itself rather than an outcome of a separate behavior program (Stokes and Baer, 1977), they began to apply experimental procedures to behaviors in settings which more closely approximated desired stimulus conditions (Gable, Hendrickson, and Strain, 1978; Jackson and Wallace, 1974; Stokes et al., 1974). As a result, various techniques have been recommended to promote generalization of training effects (Kazdin and Bootzin, 1972;

Langone and Westling, 1979; Stokes and Baer, 1977; Wehman *et al*., 1977). These recommendations include identification of relevant target skills, careful planning of treatment environments, introducing naturally reinforcing contingencies, using a variety of teachers, and actually training for generalization.

Identification of relevant target skills.

Selection of target skills which are meaningful to the subject and relevant to everyday activities is one method of influencing generalized performance. Mastery of skills which serve to improve the quality of daily existence would seem likely to result in the initiation of the skills in settings outside of experimental conditions.

Gable *et al*. (1978) selected social interaction behaviors for target skills only after systematically observing the children in free play and identifying behaviors that generally elicited positive social interactions. The resulting generalization of these carefully chosen skills, in the absence of prompts and reinforcement, was partly attributed to this concern for selection of relevant target behaviors. Selection of culturally normative play skills as target skills in PREP, a play program for young mentally retarded children (Watkinson and Wall, 1980), was based on similar attention to children in free play.

Careful planning of treatment environments.

By structuring the treatment environment to closely approximate desired generalization settings, transfer of learning may be more likely to occur. Special considerations might include selection of equipment, selection of teachers, number of children, and skill level of other chil-

dren. This planning may include teaching other children the experimental skills so that they provide realistic models.

Although the modeling skills of mentally retarded children are generally inferior to those of their non-retarded peers, instances of incidental learning have been formally and informally observed. Cooke and Appoloni (1976), upon equally dividing a group of young learning disabled children and training only one group, observed improvement in appropriate social behaviors for both the trained and untrained children when they were placed in an integrated play situation. Whitman, Mercerion, and Caponigri (1970) noted that when reinforcement programs were applied to the social behaviors of experimental subjects, the quality of social interactions among non-experimental severely mentally retarded children within the room also improved. So it is conceivable that modeling and incidental learning can occur with mentally retarded persons and if they occur they may aid generalization.

Introduction of naturally reinforcing contingencies.

Basic to the success of any instructional program are well defined instructional strategies, clear and concise prompting techniques, and systematic fading of prompts and reinforcement schedules (Wehman *et al.*, 1977). Introduction of naturally reinforcing contingencies, as early in a program as possible, is another recommendation for generalization programming (Stokes and Baer, 1977). By administering reinforcers that are likely to occur naturally in the subject's environment, generalization of behaviors is more likely to occur. This concept lacks clear empirical evidence, suggesting the need to look more seriously at this possibility.

Guralnick and Kravik (1973) found that behavior rates improved maximally with edible reinforcement yet behaviors plunged towards the pretreatment baseline when the candy reinforcement was removed. Pierce and Garland (1977) evaluated four training procedures including instructions, praise, goal setting, and materialistic reinforcement. They concluded that material reinforcers were most powerful in improving the gross motor skills of six physically handicapped mentally retarded persons, but the skills did not generalize. It is interesting to note that while edible and material reinforcement is most effective in changing behavior under initial stimulus conditions, these studies demonstrate that they may be the least effective in producing generalized effects.

Research has clearly demonstrated the powerful role that teacher contingent attention and social reinforcement may play in improving motor and social skills (Buell, Stoddard, Harris, and Baer, 1968; Hardiman, Goetz, Reuter, and LeBlanc, 1975; Johnston, Kelley, Harris, and Wolfs, 1966; Wehman and Marchant, 1978). Yet in all of these studies, a radical change in behavior was observed when social reinforcement was removed. Newly acquired behaviors did not generalize into situations where they were not reinforced, even immediately following treatment. Although behaviors were not maintained during a return to baseline phase, performance during this phase showed improvement over the original baseline scores in three of the four studies (Buell et al., 1968; Hardiman et al., 1975; Johnston et al., 1966).

Bateman (1975) investigated the possibility of using the Premack principle to increase the frequency of desired responses. Opportunity in a preferred activity was made contingent on time spent in a less

preferred activity. This treatment was successful in modifying behaviors during treatment. However, both subjects returned to original baseline performance during the return to baseline phase.

Three different methods of treatment have been shown to produce desired change in behavior during treatment. Although these treatments did not result in maintained behavior change, behaviors treated with social reinforcement remained slightly above original baseline scores even months after treatment (Hardiman et al., 1975; Johnston et al., 1966).

Organization of teaching assignments.

Evidence suggests that programming by more than one teacher can influence the generalization of skills. Stokes et al. (1974) found that while very little generalization of a handwaving greeting to people other than the teacher was observed, generalization was significantly enhanced when a second teacher was introduced. The optimal number of teachers for programming is undetermined, but Stokes and Baer (1977) suggest that two teachers may be best.

Direct generalization training.

The previously mentioned recommendations suggest methods of planning a training program that will encourage generalization. A final suggestion is simply to train for generalization. This may be accomplished by gradually spending more and more intervention time in the generalized setting and by including treatment in the desired setting, once learning becomes apparent in experimental conditions. In other words, experimental conditions are applied to the behaviors in the desired settings. Practically speaking, this may require additional patience on the part

of the teacher. However, teachers should remember that young mentally retarded children are dependent on teacher assistance until they reach independence in all settings where the skill is needed.

Although limited, there is data to support a trend in the direction of training for generalization. Gable et al. (1978) found that target behaviors declined abruptly when prompts and reinforcement were removed. However, these behaviors increased significantly during a generalization training phase in which reinforcement was applied to all target behaviors when they were initiated. In addition, a study with a fifteen year old mentally retarded girl by Jackson and Wallace (1974) produced stimulus generalization in a new classroom only after applying treatment conditions in that classroom.

These suggestions have been gleaned from the applied literature. Although "train and hope" procedures as labeled by Stokes and Baer (1977) inundate the journals and in fact often claim to demonstrate generalized effects (Stokes and Baer, 1977), researchers have begun to realize that it is unrealistic to think of generalization as a natural outcome of every behavioral program.

PREP Programming for Generalization

Organization of the PREP Program and the evolution of strategies into current PREP curriculum materials (Watkinson and Wall, 1980) have been greatly influenced by proposed methods of maximizing skill generalization. A number of strategies have been adopted in an attempt to assist children to perform independently as early as possible and to maintain newly acquired skills in the absence of prompts and reinforce-

ment.

The selection of forty PREP target skills has been based on the observation of culturally normative play behavior. Play skills which seem most meaningful to the free play behavior of young children have been identified and task analyzed for use in the program (Watkinson and Wall, 1980).

The treatment environment, the PREP room, has been structured to provide a stimulating play area including equipment typically found in both indoor and outdoor playgrounds. In addition to the climbing apparatus, slides, and trampoline in the room, tricycles, wagons, scooters, balls, hoops, and other small play equipment are available to the children during free play.

The teachers in the program are trained to use the detailed prompting continuum which provides them with a means of gradually fading their assistance during individualized instruction. When the prompting continuum is used properly, there appears to be an inverse relationship between teacher assistance and student independence. That is, as the teacher reduces her assistance, the amount of student independence increases.

Identified reinforcement and information feedback strategies in the PREP Program promote fading to naturally occurring contingencies as early as possible. A continuum of reinforcement behaviors ranging from manipulative feedback to a variety of environmental feedback have been delineated (Watkinson, Jordan, and Wall, 1980). Teacher behaviors in the phase of instruction following a response by the child may involve delaying reinforcement and feedback, or fading reinforcement to an intermittent schedule.

Instruction in certain PREP skills seems to be reflected in the speed with which a teacher can effectively fade her assistance. While preschool children find all the PREP skills enjoyable enough to use them frequently in free play, some of the skills seem to be naturally reinforced more strongly by such feelings as speed and vertigo. Consequently, they are acquired more quickly than others (Watkinson, 1977). So not only does the PREP Program provide a systematic method for fading teacher assistance, but the very nature of many of the target skills seems to promote fading rather quickly to situations where the child is prompted by the play environment.

The PREP Program recommends that teachers teach skills on a variety of similar but different apparatus whenever possible. For example, a variety of inclines including two large playground slides are typically used in teaching the skill of sliding down a slide. Varying the stimulus conditions in this manner may help to maximize generalization.

Behavior Categories Which Describe Generalization

It has been established that the refinement of principles and procedures which can explain generalization is an area requiring attention. While a researcher may not find established criteria to evaluate generalization in a study, it may be possible to gather enough information from the literature to become confident that he has chosen good measurement and evaluation standards. This is especially true if the purpose of the study is to look at a class of behaviors such as social interaction skills.

However, if the behavior of interest is a discrete motor skill, the literature does not provide many examples of measurement methods. Gener-

alization of motor skills has typically been evaluated in terms of target behaviors and approximations to the target behaviors. One example of a study describing the generalization of a task analyzed play skill receiving instruction measured the climbing behavior of a subject whenever he made physical contact with the climbing apparatus (Hall and Broden, 1967). Similarly, Johnson et al. (1966) defined climbing-frame behavior as physical contact with the frame or with auxiliary parts attached to it. Flavell (1973) also included approximations to performance of the target skill in her evaluation of play. Are these approximations adequate for describing generalization?

Hardiman et al. (1975) included four levels of performance in an attempt to develop a reliable recording system to assess children's motor skill performance. The four different levels included proximity to equipment, touching equipment, unskilled performance, and skilled performance. Behavioral definitions were identified to clearly describe skilled and unskilled performance. Percentage scores for the proximity and touching categories were eliminated in the final evaluation of performance. The skilled and unskilled performance scores were graphed separately and combined to represent an active participation score.

Given the promise of category observation instruments in the evaluation of critical variables in learning (Martin, 1976), use of similar instruments in the description of behaviors important to the evaluation of generalization is an interesting possibility. The use of a detailed continuum of behaviors which approximate the skilled performance of a target skill in free play may provide information which, to this point, has not been provided elsewhere.

Measurement of Stimulus Generalization

The meaning of stimulus generalization is to a large extent dependent upon where generalization is measured and when it is measured. A variety of different situations can be considered as generalized settings.

The degree of similarity between a generalized setting and the experimental setting is crucial to the interpretation of stimulus generalization data. A more realistic evaluation of the transfer of learning is obtained when the discrepancy between situations is great. For example, free play behavior can be recorded in the same room where children receive instruction, or in a different room. Observation in a different play setting, as Jensen (1979) chose, is the more conservative method of evaluating generalization. The actual physical environment in which stimulus generalization is measured depends largely on the purpose of the study.

Generalization can be measured at a variety of times in relation to treatment intervention. Treatment sessions can be divided into two parts with experimental conditions applied during the first half and observations of generalization in free play made immediately following. This is a rather lenient method of generalization data collection because of the proximity in time between treatment and the measurement of stimulus generalization. Yet, generalization is very often reported this way in the literature (Cooke and Appoloni, 1976; Gable <u>et al</u>., 1978; Guralnick and Kravik, 1973).

A more conservative measure regards the identified target behavior at some time during the day or week, other than when experimental conditions were in effect. For example, generalization data can be collected in the afternoon after a treatment session has been held in the mor-

ning, or at random times throughout each day (Walker and Buckley,1972; Whitman et al., 1970).

An even more cautious measurement of the practical value of an instructional program can be made by looking at the maintenance of behaviors. Response maintenance or long term generalization (Cuvo, Leaf, and Burakove, 1978), refers to the occurrence of behaviors in everyday settings after treatment intervention has ended. Behavior changes which are of practical significance to the subject may be measured best this way. Practical problems, such as time and money, often make long term follow up measurements difficult or impossible.

Quite obviously, a more conservative estimate of the transfer of learning to typical everyday situations is made when a greater time span elapses between treatment and generalization measurement. Cuvo et al. (1978) made post checks over a two week period to illustrate skill maintenance. Walker et al. (1975) collected follow up data during the first four months of the school year following treatment.

Regardless of when data is collected, values obtained during a baseline phase and a generalization phase are compared. The researcher must evaluate the data to determine whether generalization did or did not occur.

Where once the critical question in generalization was the identification of procedures for facilitating generalization (Wehman et al., 1977), the current question concerns the need to formally quantify generalization criteria. Until some type of "quality control" is identified to evaluate the occurrence or non-occurrence of generalized behavior, the variety of evaluation standards reported in the literature is not

likely to change. The issue presently requiring attention addresses the question of how much generalization is enough to be practically significant to the subject. Included in this quality control should be some type of standard definitions for describing generalized behavior. Although this would be difficult because behaviors are often specific to the individual study, it may be an important consideration.

Category Observation Instrumentation

The observation of human behavior in natural settings is an important form of research. Martin (1976) suggests that direct observation techniques can provide more accurate information and understanding about teaching and learning than any other single source. Unfortunately, only recently have direct observation techniques and single subject designs begun to be recognized as credible forms of experimentation (Kratochwill, 1978). Given the promise of category observation instruments in the evaluation of critical variables in learning, it is exceedingly important that they be developed and utilized in ways which will yield valid and reliable information (Martins, 1976).

Direct observation research is based on collections of descriptive categories. The descriptive categories chosen are a function of the behavior or behaviors investigated (Connolly, 1973). Previously designed category observation instruments often prove inappropriate for describing a set of behaviors under study and a unique set of categories must be developed.

The generation of descriptive categories is an important and difficult task. Upon deciding which aspects of behavior should be recorded,

the behaviors must be broken down into clearly observable and measurable components. It is essential to write definitions which are clearly objective since vague and subjective definitions become a source of error in obtaining accurate and objective data (Hawkins and Dotson, 1975).

A proposed instrument should then be tested to ensure that it adequately covers the range of behaviors being investigated. Flanders (1966) makes recommendations to assure this representation. First, the categories should be designed to include all possible events. Secondly, categories should be defined in such a way that any single event can be recorded in only one category; that is, the categories must be mutually exclusive. Failure to meet these requirements results in data which is not representative of all events or biased frequencies when data in categories are compared.

Rules cannot be set to limit the number of categories included in an instrument. There is however, a limit as to how much information an observer can process and record in the selected time interval. Inevitably, a trade-off must be made. A large number of categories provides for specificity and fine distinctions but creates problems in regards to accurate measurement (Kazdin, 1977) and inter observer agreement. An instrument with a small number of categories provides a more general description of behavior patterns, making recording easier but reducing the discriminitive power of the instrument (Connolly, 1973).

Martin (1976) outlined several criteria for use in the evaluation of observational systems. First, the instrument must be objective; that is, it must include operationally defined categories. The behaviors must be relevant and have a likelihood of occurring in the setting being

studied. Third, every effort should be made to reduce the number of categories and simplify the recording process as much as possible. An efficient instrument reduces the number of categories so that behaviors are tallied across all categories, eliminating seldom employed categories. The last two properties of validity and reliability are requirements of any measurement device. Reliability implies similar coding of behaviors by two independent observers and a valid instrument differentiates between observed behaviors, rather than between impressions of what was actually observed by different observers. A sound category instrument meets these requirements of objectivity, relevance, parsimony, efficiency, reliablity, and validity (Martin, 1976).

Sampling Techniques

Once a category observation instrument has been established, a method of collecting data must be selected. Basically there are two choices: continuous or sampling techniques (Connolly, 1973). In continuous procedures, data are collected during the entire period that the behavior under study had the potential to occur. Quite clearly, continuous data collection is a very thorough method, although it is often not the most practical. If the duration of a study is lengthy, a sampling method is more practical, since data need be collected only during part of the total time. Carefully chosen and employed sampling procedures produce a good representation of the total picture.

In selecting an appropriate sampling method, the researcher must decide between event sampling and time sampling (Connolly, 1973). Event sampling techniques provide for precise measures of behavior in terms

of frequency and duration over a specified length of time (Hutt and Hutt, 1974). The observer records an occurrence every time a defined behavior occurs. The resultant data is expressed as frequency, frequency/time, or as duration when time measures are made in addition to the frequency count.

In time sampling, observations are made in a series of short intervals providing a representative sample of the behaviors being studied. The length of the interval is usually determined by the frequency and duration of the defined behaviors. Short intervals are used to capture infrequent and discrete behaviors while longer intervals can be used if behaviors are of greater length or frequency. The samples can be discontinuous or sequential. By using sequential samples, a long observation period can be broken down into short intervals the duration of which may last from two to thirty seconds. Sequential time sampling is difficult to employ but if used, produces results which closely approximate continuous recording data (Connolly, 1973).

An interval time sampling system requires that the observer record given responses when they occur within the interval (Sulzer-Azaroff and Mayer, 1977). When the behavior is emitted for the entire interval to be scored, it is whole interval time sampling. Partial interval time sampling requires that the behavior occur at some time during the interval. A third method, momentary time sampling, scores behaviors that are occurring when the interval ends. The purpose of the measurement determines which system should be used.

Arrington (1939) cleverly describes interval time sampling as a form of controlled observation in which the behavior, the method of

recording, and the manner of selecting the behavior to be observed are subject to control rather than the situation in which the observations are made.

Selection of the appropriate sampling technique and the meaning of the data obtained depends largely on the duration and frequency of the behaviors being studied and the anticipated total observation time.

Inter Observer Agreement

In using direct observation as a tool for measuring behaviors in natural settings, it is necessary to know that the behaviors are being recorded consistently. In the field of applied behavior analysis, consistency is expressed in terms of inter observer reliability. This reliability refers to the agreement attained by independent observers when coding the same sequence of behaviors (Connolly, 1973). Two or more observers code a set of behaviors and a reliability coefficient is calculated employing one of several methods.

Typically in determining observer agreement, one observer's data serves as a criterion and the number of agreements and disagreements is totaled with the assumption that the criterion data are without error in reflecting the actual behavior of the subject. Kazdin (1977) challenges the idea that agreement calculated in this manner is really a measure of "true" performance. Although the two are related, accuracy and agreement are not the same (Kazdin, 1977). The possibility of two observers agreeing to an incorrect observation exists. Bijou and Peterson (1969) suggest that using a third observer on occasion may check observer accuracy. Kazdin (1977) has left the dilemna unanswered,

recommending that an agreement coefficient of 70% - 80% is considered satisfactory for reflecting accuracy, while at the same time proposing that any quantitative criterion may be senseless.

The degree of agreement or reliablity is influenced by a variety of factors, including: (1) complexity of the coding instrument, (2) specificity of the definitions, (3) pattern of behaviors coded over sessions, (4) frequency of coded behaviors, (5) selection of observers, (6) observer training, (7) observation conditions, (8) length of observation period, (9) frequency of reliability checks, and (10) the method used to calculate the reliability coefficient (Bijou and Peterson, 1969; Hawkins and Dotson, 1975; Johnson and Bolstad, 1973; Kazdin, 1977). Recommendations have been made to minimize these possible sources of error in observing human behavior (Bijou and Peterson, 1969; Kazdin, 1977) to preserve the meaning of reliabilty estimates.

It is up to the investigator and the purposes of the study to decide which agreement method to use. However, it is essential that agreement data be calculated on the score identified as the dependent variable (Johnson and Bolstad, 1973).

One method of calculating observer agreement called scored interval agreement (Hawkins and Dotson, 1975), employs the following formula:

$$\frac{a}{a + d} \times 100 = \% \text{ agreement}$$

and considers only intervals in which both observers coded the presence of a particular behavior as agreements. Hawkins and Dotson (1975) differentiate this method from the interval-by-interval method where intervals in which both observers record the absence of the particular

behavior are also counted as agreements. These same investigators have reported that interval-by-interval scores are often grossly unreliable because they are highly subject to influence by the rate of the behaviors being studied. Although there are limitations to the scored interval method, it is an improvement over interval-by-interval agreement scores (Hawkins and Dotson, 1975).

Single Subject Design

The aim of scientific investigation is to demonstrate a relationship between changes in behavior and changes in treatment. When eliminating the numerous alternative explanations for a cause-and-effect relationship, it does not suffice to show that intervention with a particular behavior was followed by an increase or decrease in the behavior. The experimenter must establish that the experimental procedures, rather than other variables in the subject's environment, were responsible for the observed increase or decrease.

This causal relationship referred to as internal validity by Campbell and Stanley (1963) is threatened by several confounding factors. These factors include history, maturation, testing, instrumentation, statistical regression, subject selection, experimental mortality, and selection-maturation interaction (Campbell and Stanley, 1963). Although every one of these may not be a threat to every experiment, the scientist must control for the variables which could produce effects confounded with the treatments under investigation. Regard for internal validity is paramount to the believibility of experimental results.

A valid conclusion that a treatment actually produced the observed

behavior changes can only be made if the data is collected under controlled conditions. Selection and application of an appropriate design helps to screen out possible influences of confounding variables and strengthens the internal validity of a study.

Traditionally, group experimental designs have been accepted as useful instruments in minimizing confounding effects and adequately representing group scores. It seems however, inappropriate to draw any conclusions about individuals on the basis of a group design. Consequently, stringent single subject designs have been developed to meet the needs of researchers interested in treatment effects on individuals (Baer et al., 1968; Bijou and Peterson, 1969; Birnbrauer, Peterson, and Solnick, 1974; Hanley, 1970).

One such single subject design is called the reversal or ABA design (Baer et al., 1968). The reversal design is based on the logic that behavior changes are contingent upon the presentation and withdrawal of treatment variables. Behaviors measured under experimental conditions (phase B) are compared to a pre-experimental baseline condition (phase A) which serves as a representation of the subject's typical behavior. A direct relationship between the experimental conditions and any observed behavior change is strengthened if the subject returns to pre-treatment behavior rates upon withdrawal of treatment. Withdrawal of treatment is referred to as a second baseline condition (A) thus, the ABA design.

Generally speaking, the return to baseline phase is included in a study as a means of demonstrating that the observed behavior change was contingent with the applied treatment. Use of an ABA design to establish experimental control, when behaviors are unlikely to return

to baseline during the second baseline phase, is not appropriate. When behaviors are learned during treatment intervention and/or come under the control of naturally occurring contingencies, behaviors are not likely to return to baseline performance as a result of terminating treatment. Therefore, instructional researchers seldom use the reversal or ABA design as their only design. However, this design is frequently used in combination with other designs when learning is likely to occur (Flavell, 1973; Hardiman et al., 1975).

Alternative designs are available for use in establishing experimental control when it is suspected that behaviors will not return to baseline rates when treatment is withdrawn. Multiple baseline designs offer an alternative (Baer et al., 1968).

One type of multiple baseline design is called the multiple baseline design across behaviors. Initially, baseline data is collected on a number of selected behaviors. Treatment is applied to one behavior while all identified behaviors continue to be monitored. When the treatment produces a change, the experimental conditions are applied to another, presumably unchanged behavior. Successive apparent causal relationships increase confidence in the treatment. Evidence that the behaviors change only when they are placed under experimental conditions strengthens the belief that this change is more than coincidental.

Selection of an appropriate experimental design is governed by the questions under investigation. Recent behavioral research often includes a combination of single subject designs (Cuvo, 1979; Flavell, 1973; Hardiman et al., 1976; Strain and Shores, 1979). Most often when more than one design is used, one design examines the questions asked

in the study while the companion design is used to demonstrate experimental control.

A study which is looking at the generalization and maintenance of instructional skills can be efficiently conducted using a combined multiple baseline design across behaviors and an ABA design, where the treatment is withdrawn in the second A phase. Use of the multiple baseline design helps to establish a cause and effect relationship between the treatment and observed generalization of behaviors. While the ABA portion of the design will not establish experimental control with learned behaviors, it is useful in the evaluation of the maintenance of behaviors following treatment.

CHAPTER III

METHODS AND PROCEDURES

Introduction

The foremost question being asked in this study was whether the individually prescribed play skills of instruction generalized into the structured free play environment. A multiple baseline design across behaviors (Baer _et al._, 1968) in combination with an ABA design (Baer _et al._, 1968), was selected as the most appropriate design to answer this question. Treatment conditions were withdrawn in the return to baseline phase of the ABA design.

Continuous videotape recordings of the children's behaviors while participating in the PREP Program were collected during every instructional session. These recordings served as raw data for the study. They were coded by the criterion observer and two other trained observers.

Subjects

The subjects in this study were eight moderately mentally retarded children who participated in the PREP Program at the University of Alberta. The children were pupils in two early education classrooms at Winnifred Stewart School for Retarded Children in Edmonton, Alberta. Their mean age at the start of the study was 86.3 months (range 63-107 months), the means for the five boys and three girls were 87.8 months and 83.7 months, respectively. The children were transported by bus to the PREP Program two mornings a week for seventy five minutes of free

play and instruction.

Each subject met the following criteria for inclusion in this study:

1. The child had a good attendance record in the program.
2. The child was without secondary physical handicaps.
3. The child was under nine years of age.

Of the eight subjects who met these criteria, three boys and three girls were diagnosed as having Down's Syndrome. The other two children have unknown etiologies.

The Teachers

The teachers in the PREP Program for the duration of this study included three graduate students (including the investigator) and two senior undergraduate students, all specializing in the area of adapted physical education. All teachers had been trained in the PREP procedures and had worked in PREP for several months prior to this study. The selection of this sample was not random, but was made on the basis of availability and good will. The two undergraduate students were paid and the two graduate students were volunteers.

These teachers underwent a further training program covering assessment, instruction, and maintenance procedures specific to this study. In addition to this training, the teachers met weekly to discuss the results of instruction and maintenance as indicated on the daily record forms and graphs.

Treatment

This study was conducted within the PREP Program at the

University of Alberta. The central feature of the PREP Program is an ordered set of task-analyzed, instructional sequences designed to facilitate the assessment, selection, and teaching of culturally normative motor skills (Watkinson and Wall, 1980). Instructional strategies provide techniques for prompting performance while fading teacher assistance.

The PREP instructional system includes a prompting continuum with the following categories: physical, visual, verbal, and environmental. The prompting model (Figure 1) suggests that teachers begin with physical assistance and fade their assistance, eventually to situations where naturally occurring contingencies in the environment prompt the child to perform. The prompting system is used in conjunction with appropriate reinforcement and information feedback to move the children through the task sequences towards the target skill (Watkinson, Jordon, and Wall, 1980).

Within a PREP session, individualized instruction is carried out while the children are engaged in play. A teaching episode involves a short interaction between a teacher and one child (Watkinson and Wall, 1980). Teachers carefully complete a daily record form immediately after each instructional episode.

PREP sessions were held each Wednesday and Friday morning during the months of May and June, 1980. Each PREP session was divided into fifty minutes of free play time with individual instruction, a five to ten minute group activity, and an additional fifteen minutes of free play with tricycles and wagons available.

The organization of the PREP playroom indicating the arrangement of available play apparatus is depicted in Diagram 1.

FADING TEACHER ASSISTANCE

INCREASING INDEPENDENCE

PHYSICAL PROMPTS
- Complete Manipulation
- Partial Manipulation
- Minimal Guidance

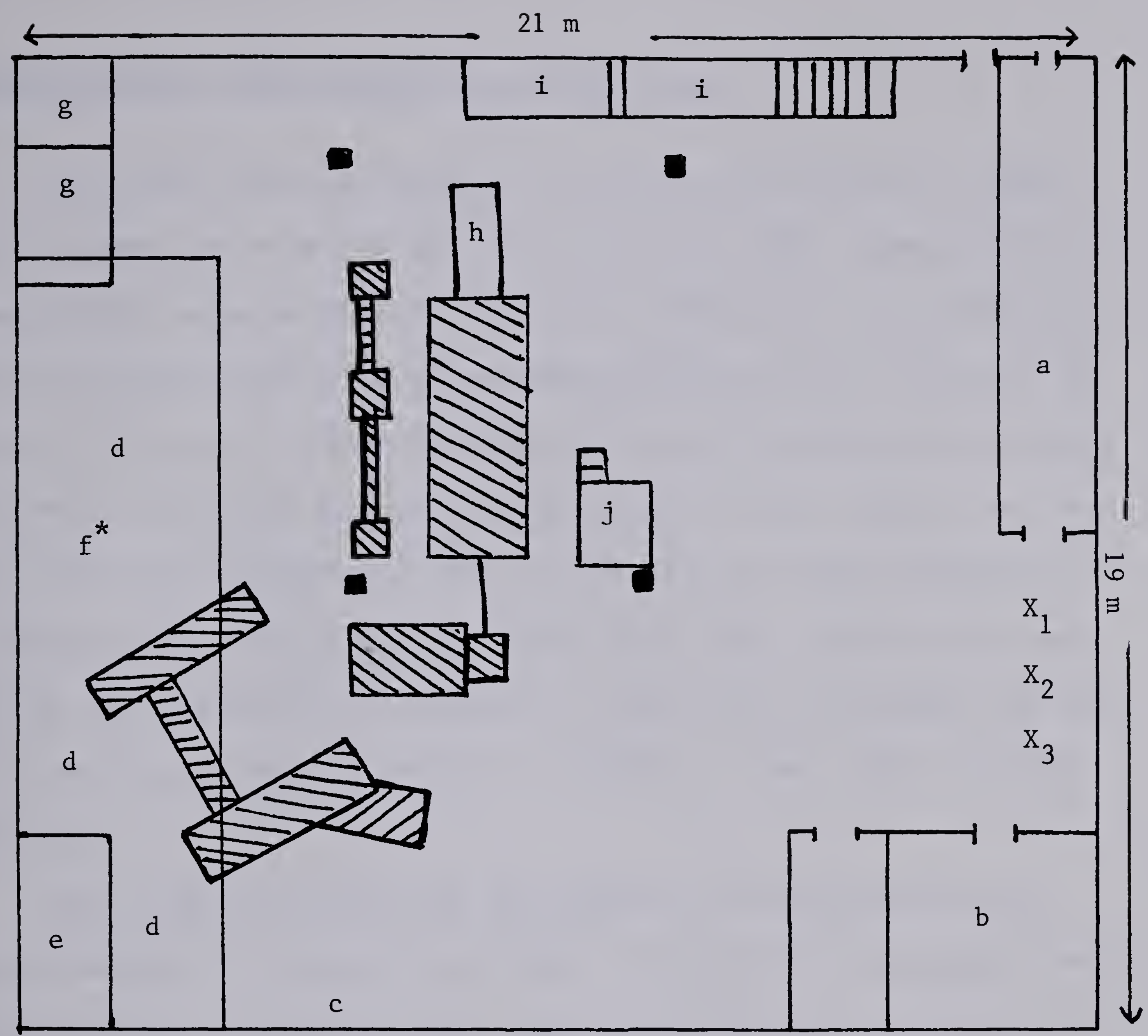
VISUAL PROMPTS
- Complete Demonstration
- Partial Demonstration
- Gesture

VERBAL PROMPTS
- Skill Cues (e.g. "bend your knees")
- Skill Mands (e.g. "jump")
- Action Words (e.g. "go")

ENVIRONMENTAL PROMPTS
- Structured Goals
- Peer Demonstration (Spontaneous)
- Initiation

Figure 1: PREP Model for Prompting Performance and Fading Teacher Assistance.

Diagram 1 : PREP ROOM



■	support pillar
▨	climbing apparatus
┤ ├	door
a -	observation room (one way mirror and TV monitor storage)
b -	office
c -	throwing target
d -	large mat (4m x 15m)
e -	trampoline
f*-	swing (suspended from ceiling)
g -	scooter ramp
h -	padded slide
i -	large slide (2m. ladder, 3m. platform, 3m. long slide)
j -	steps and platform
X -	camera

Modifications of PREP Strategies for this Study

Six skills were prescribed for each child from the list of PREP task-analyzed motor skills (Watkinson and Wall, 1980). These skills were thoughtfully selected on the basis of the PREP skill prescription guidelines and three hours of free play observation noted just prior to the start of this study. Each prescribed skill had been initiated two times or less by the child during these three hours of play time and furthermore, had been assessed according to the PREP assessment guidelines at a level of independence. In other words, the child could perform each of the skills prescribed for him/her at some task step without any physical assistance from the teacher but seldom, if ever, chose to during free play.

Two of the six skills were prescribed for instruction and four were prescribed for maintenance. Skills prescribed for maintenance were assessed at all levels including target skill level while skills prescribed for instruction were initially assessed at a level below target skill.

The prescribed skills were divided into two groups so that each child had one instructional skill and two maintenance skills for the first eight sessions of instruction, and a new instructional skill and two new maintenance skills for the final six sessions. A list of subjects and prescribed skills can be found in Appendix A.

Skills prescribed for instruction were taught in accordance with PREP instructional strategies discussed previously. Three to ten minute instructional episodes with individual children were interjected into free play time. Teachers were encouraged to fit three instructional epi-

sodes into each session.

Skills prescribed for maintenance were taught according to the maintenance strategies identified by the investigator. These strategies included usage of the PREP prompting continuum, with the exception of the physical prompting categories. Because the children could perform each of the skills prescribed for maintenance without any physical assistance, maintenance strategies included walking away from a child when he did not respond to a complete demonstration by the teacher. In other words, the child was not given physical assistance because it had been established that he could perform the skill independently at will. The prompting system used for maintenance skills included environmental, verbal, and visual prompts. Teachers were encouraged to fit two maintenance episodes for each skill into every session.

Teachers were required to complete a daily record form for each child, immediately following each instructional and maintenance episode. Daily graphs for each child were kept up to date by the investigator.

Teaching assignments were arranged so that each teacher interacted with every subject during the eight week program. These interactions could be either instructional episodes, maintenance episodes, or both. It was felt that this type of arrangement might make learning more pleasant for teachers and students, and would foster the generalization of skills to other stimulus conditions.

Teaching assignments were planned this way to eliminate the problem of one teacher repeatedly interrupting a child's free play. Based upon the data collected, it was found that the children in this study actually spent an average of 35% of their free play time in teacher inter-

action. This indicates that a considerable amount of the children's free play time may have been interrupted by instruction. It is quite easily understood that a child would become impatient, and perhaps unwilling to perform a skill, if one teacher repeatedly interrupted the child on seven occasions during each session. It was felt that the children's performances might be maximized by dividing these interruptions among three different teachers.

Time and Duration of the Study

The data for the pilot study was collected during the month of March, 1980.

The data in this study was collected during the months of May and June, 1980.

Instrumentation

Development of the instrument.

A pilot study was conducted in March, 1980. The purpose of this study was to examine the possibility of using VTR equipment to record activity in the entire PREP playroom, and to devise a category observation instrument appropriate to describe the generalization of skills receiving instruction into a free play setting.

A continuum of behavior categories was proposed after repeated observation of the free play behavior of moderately mentally retarded children. The proposed instrument included the following categories: Attends to Equipment, Touches Equipment, Attends to Another Child on the Equipment, Physical Contact with Equipment, Inappropriate Object Manipulation,

Object Manipulation, Performance Below Target Skill Level, Performance At Target Skill Level, and Performance Above Target Skill Level.

Four children who were participating in the PREP Program were selected as subjects. Activity during sixty minutes of the PREP Program was filmed on five separate days. Each tape was coded in four fourteen minute segments.

All behaviors observed fit into one of the nine categories. Some categories had a greater occurrence recorded than others. It was concluded that the behavior categories were relevant; that is, they adequately described the behaviors under study. In addition, an optimum filming arrangement was determined.

Revised instrument used in this study.

A revised behavior category instrument was designed for use in this study. Changes to the instrument used in the pilot study became necessary when sampling procedures were changed. Partial time sampling was selected for this study, whereas event recording was used in the pilot study.

The revised behavior category observation instrument included the following categories.

Touches (1)

Some part of the body comes in contact with the apparatus of instruction without regard for the particular piece of equipment. There is no manipulation of the equipment.

Example: A child sits on the floor and places one hand on a stationary scooter.

Attends (2)

Active interest in a piece of equipment of instruction or another child's performance of a skill of instruction. There is apparent intention on the part of the observer. This does not require mutual participation.

Example: A child sits on a bench and watches another child ride around the room on a scooter.

Physical Contact (3)

Direct contact is made with regard for the piece of equipment of instruction, without any manipulation of the apparatus.

Example: A child sits on a scooter on the floor but does not move.

Inappropriate Object Manipulation (4)

Physical contact with a piece of equipment of instruction with purposeful manipulation. More than maintaining physical contact. The manipulation is negative in nature and directed towards another child.

Example: A child pulls a scooter out from under another child.

Object Manipulation (5)

Physical contact with apparatus of instruction with purposeful manipulation which is more than maintaining physical contact. This includes behaviors where the child appears to be preparing to perform a prescribed skill, performs a skill related to a prescribed skill, and any other manipulation of equipment of instruction other than a performance below, at, or above the last level of instruction. This also may include assisting someone else with a skill a child is being taught.

Examples: A child is being taught to ride a scooter down an incline on his tummy; any scooter skill on the floor.

Carrying a scooter while running up the incline.

Sitting at the top of the incline and giving someone else an initial push as he rides down the incline with his tummy on a scooter.

<u>Performance Below Last Level of Instruction/Maintenance/Assessment</u> (6)

Performance of a purposeful, appropriate skill of instruction, below the task step at which a child last received instruction or maintenance, or was assessed at prior to instruction.

Example: Child is being taught to sit on a scooter and ride it down the incline with hands and feet off the floor (task 3). He is seen riding down the incline dragging his feet (task 2).

<u>Performance At Last Level of Instruction/Maintenance/Assessment</u> (7)

Performance of a skill on apparatus of instruction at the task step at which a child last received instruction or maintenance, or the assessed level prior to instruction.

Example: Child is being taught to ride a scooter on his seat by pushing with both feet (task 2) and he is seen pushing with both feet simultaneously.

<u>Performance Above Last Level of Instruction/Maintenance/Assessment</u> (8)

Performance of a skill at a task step above the last step receiving instruction or maintenance; or if the last instructed level is at target skill level, the child uses the equipment more skillfully, efficiently, or socially than the target skill as defined.

Examples: Child lies down on two scooters and rides down the incline.

Child is being taught to ride a scooter on his seat by pushing with both feet (task 2) and is seen pushing with alternate feet (task 3).

<u>Playful</u> (9)

Any purposeful activity other than one of the prescribed skills or a skill using the same equipment as a prescribed skill.

Examples: Walking around the playroom.

Sitting and watching someone swing on a swing.

Riding a tricycle when tricycle riding is not a prescribed skill.

<u>Non-Playful</u> (10)

Sitting alone without interest in any play equipment or any other child, or a negative behavior which is unrelated to the prescribed skills.

Example: Hitting another child with a hockey stick, when hockey skills have not been prescribed.

<u>Unobservable</u> (11)

The child is out of view for more than half of the three second interval.

<u>Teacher Interaction</u> (12)

Any definite interaction between a child and teacher. Teacher contingent attention may be included in this category.

Examples: Teacher walks over to a child and reprimands him.

Teacher is conducting an instructional episode.

Teacher and child are playing hockey together.

Categories 1 through 8 were arranged with the assumption that the behaviors increase in sophistication as they approach category 8. Therefore, performance above target skill level was considered to be more

sophisticated than performance at target skill level, which was considered to be more sophisticated than performance below target skill level.

Mutual exclusiveness was insured by specifying that the most sophisticated behavior was recorded if a behavior fit into more than one category. Therefore, behaviors coded in category 6, for example, could include behaviors in categories 1 through 5.

Categories 6, 7, and 8 defined the level of skill performance in relation to the last task step receiving attention. The task steps of each target skill are detailed in the PREP manual (Watkinson and Wall, 1980). All skills were coded in terms of the highest level of instruction or maintenance attained during that particular session, or if the skill had not yet received instruction, it was coded in terms of the initial assessment level. In other words, behaviors coded in categories 6, 7, and 8 during the baseline phase were coded in relation to the highest task step which the child performed independent of physical assistance during skill assessment.

Categories 9 though 12 were defined to ensure total inclusiveness and guarantee active recording at all times.

Coding rules were established in a further effort to meet the requirements of total inclusiveness and mutual exclusiveness. In addition to these general coding rules, skill specific coding rules were established to maximize observer consistency. These coding rules can be found in Appendix B.

Equipment

Three SONY CVC 2100A black and white cameras, two with SONY 12.5-75 mm zoom lenses and one with a SONY 16-64 mm zoom lens, were used for

filming purposes. The cameras were attached to tripods and placed on top of small tables approximately two feet high. Optimal placement (See Appendix C) was determined in a pilot study (Terry, 1980) to permit film coverage of nearly all free play space in the PREP playroom.

Three SONY Solidstate 3/4" video cassette recorders, model VO-1800, in conjunction with three Electrohome television monitors were used to record the activity. In order to minimize subject reactivity, the three recording packages were stored in the adjacent storage room during all filming.

Data Collection

Activity during the first sixty minutes of each PREP session was recorded. All three cameras recorded simultaneously over the sixty minute period, each capturing activity in a different area of the playroom. At the start of each day's recording, a large sheet of white paper was held up in front of all three cameras and then taken away, to facilitate synchronized replay of the tapes for coding.

Coding Procedures

A partial time sampling method was used whereby each 42 minute observation period was divided into six 7 minute segments, each of which was further divided into three second intervals. "Bleeps" were electronically recorded and played back on a cassette tape recorder indicating the end of each interval. Data were collected continuously with each observer looking and coding within each 3 second interval. Three second intervals were used to allow calculation of interval proportions for categories and to compute inter-observer agreement scores.

Coding rules were delineated to ensure that the instrument was mutually exclusive for each identified skill. However, since six skills were identified for each subject, more than one behavior could be recorded in a three second interval. For example, if a child was being taught to jump on a trampoline and throw a ball, and was seen jumping on the trampoline with a ball in his hand, both behaviors were recorded. In accordance with the coding rules and the definition of a partial time interval sampling system (Sulzer-Azaroff and Mayer, 1977), coded behaviors were observed during at least part of the 3 second interval.

All data were collected on prepared coding sheets. Category symbols (see Appendix D) were used in each interval. In addition, the observer recorded the specific skill when she first started coding a generalized behavior.

Observer Training

Observer training sessions with the investigator and two observers were held prior to and during the early stages of filming. Initial training consisted of memorizing the behavior categories and coding symbols. Proficiency in this was fostered by the review of recorded tapes with discussion of observed play behaviors. Scores of 88% for Observer A and 85% for Observer B on a training quiz (Appendix E) met the predetermined criterion of 80%.

In the next training session, coding rules were explained and discussed. Detailed coding rules can be found in Appendix B.

Training continued with simultaneous verbal coding of short segments at three second intervals, coding of seven minute segments using the de-

signed coding sheet, and finally simultaneous coding of an unfamiliar seven minute segment to the predetermined level of accuracy (r=.80). The investigator's data served as criterion standard, and the agreement test produced reliability coefficients of .86 for Observer A and .88 for Observer B, when the scored interval method was utilized across all categories.

Observer training was conducted over four sessions and lasted approximately ten hours. In addition, short refresher sessions were interspersed throughout the six weeks of coding. These sessions involved discussing current coding problems, writing additional coding rules, and reviewing behavior definitions and coding procedures.

Inter Observer Agreement

The coding of data in this study was completed over a six week period. All three observers were present at the majority of coding sessions. It became necessary at certain times to hold coding sessions with only two observers. The criterion observer attended every coding session.

Each of the three observers coded the free play behavior of individual children during a coding session. In addition, the criterion observer simultaneously coded 17% to 25% of all data coded by the other two observers. This data overlap was collected for the purpose of establishing observer accuracy.

Treatment of Data

Two different graphic representations were presented in this study in the attempt to answer the questions proposed. Firstly, graphs were

constructed for each subject with the percentage of occurrence of behaviors in the separate categories serving as dependent variables. In addition to the separate categories, two different composite scores were represented graphically. Secondly, behaviors related to the specific skills receiving treatment were graphed for each subject, with the dependent variable identified as the percentage of occurrence of behaviors initiated at a skilled level of performance. Skilled performance was defined as performance of a skill receiving treatment at a level below, at, or above the last task step receiving instruction or maintenance.

All behaviors were presented as a percentage score. Percentage scores were calculated according to the following formula: (number of intervals observed/ number of intervals of free play) x 100% = percentage of occurrence in free play.

One of the eight original subjects was deleted from the analyses due to frequent absences. Consequently, the conclusions of this study were based on the results of the performances of seven subjects.

Prescription methods specific to this study influenced the method of analyses selected. That is, the children were prescribed six skills individually on the basis of their free play performance and apparent interests, so that each subject received treatment in different skills. Therefore, the percentage scores for each subject actually represented generalized performance in different skills. As a result, the dependent variable did not measure exactly the same behavior for Subject 1 as it did for any other subject. Use of a group statistic could not take this into account and therefore seemed inappropriate for use in this study.

While arguments can be posed to defend both statistical and gra-

phic analysis of data, careful graphic presentation has gained respect in recent years. Concern for evaluation in terms of statistical properties such as baseline stability, data variability, trend changes, and level changes has helped to strengthen the believibility of graphically analyzed data. Additionally, it seems logical to evaluate the success or failure of programs with the goal of changing individual behaviors, in terms of individuals.

The first set of graphs were visually analyzed with respect to data variability, trend change, and level change. The second and third sets of graphs were visually analyzed with respect to criteria specified by this investigator, further elaborated upon in Chapter 5.

CHAPTER IV

RESULTS

Introduction

The ultimate purpose of this study was to evaluate the generalization of motor skills in the PREP Program. Because a category observation instrument was used to accomplish this task, it was necessary to establish the accuracy of the instrument. For this reason, inter observer agreement scores were calculated. These calculations are presented in the first part of this chapter.

Because the instrument included a range of behaviors which could be considered evidence of generalization, an assessment of the sensitivity of the separate categories to changes observed seemed a prerequisite to the evaluation of generalization. Performances in the individual categories were graphed to aid the selection of an appropriate dependent variable to describe generalization. These graphs are presented in Figures 2 through 10c in this chapter.

Individual subject performance graphs were then constructed utilizing the selected dependent variable. These graphs, included in this chapter, seemed to indicate that some skills may have a greater tendency to generalize than others. Individual subject graphs for skills prescribed to three or more children were grouped together to allow an examination of this possibility. These graphs are included in the final section of this chapter.

Inter Observer Agreement

Inter observer agreement scores were calculated using the formula for scored interval agreement (Hawkins and Dotson, 1975). During the sixteen coding sessions, a total of fourteen reliability checks were made between the criterion observer and observer A, and fifteen checks with observer B. The criterion observer simultaneously coded 17% to 25% of all data coded by the observers in each coding session. Reliability coefficients were calculated on this overlap.

Inter observer agreement scores averaged 90% and 94% for observers A and B, respectively. Individual category reliability scores ranged form 50% to 100%. Table 1 summarizes these data calculated for each observer in each behavior category.

Generalized Behavior as Described by a Category Instrument

While the category observation instrument used in this study was designed such that any one of the categories could serve as a dependent variable to describe generalization, it was suspected that some behavior categories would be more appropriate in measuring skilled generalization than others. To answer this question and identify a conservative and realistic measurement of generalized training effects, behaviors were graphed over sessions for each separate category, with the exception of the Inappropriate Object Manipulation category. This category was excluded due to extremely infrequent occurrence. In addition to the seven remaining categories, two different composite scores were graphed. First, the scores in the three most sophisticated categories, Performance Below Last Level of Instruction, Performance At Last Level of Instruc-

Table I : Inter Observer Agreement Scores For Each Behavioral Category Over All Coding Sessions

Behavior Category	Observer A		Observer B	
Playful	90%		97%	
Non - Playful	95%		80%	
Teacher Interaction	93%		94%	
Unobservable	94%		94%	
Touches	50%		73%	
Attends	59%		76%	
Physical Contact	70%		80%	
Inappropriate Object Manipulation	-		-	
Object Manipulation	92%		91%	
Performance Below Last Level	86%		78%	
Performance At Last Level	96%		79%	
Performance Above Last Level	-		100%	
	Mean	Range	Mean	Range
All Categories Together	90%	78-96	94%	88-99
All Generalized Behaviors	85%	64-100	82%	62-100

tion, and Performance Above Last Level of Instruction, were summed and graphed as a skilled performance score. Also, the scores in the seven categories evaluated separately were summed to represent a composite score. Graphs were constructed for each subject with PREP sessions identified as the independent variable. The percentage of occurrence of the defined category served as dependent variable.

Percentage scores were calculated for each of the six prescribed skills. For the purposes of this evaluation of category efficacy, percentage scores for the three skills treated in the treatment A phase were summed into one score representing the percentage of free play time spent performing treatment A skills in each session. Similarly, the three treatment B scores were collapsed into a second daily score.

Performance graphs for the categories of Touch, Attend, Physical Contact, Object Manipulation, Performance Below Last Level of Instruction, Performance At Last Level of Instruction, Performance Above Last Level of Instruction, the skilled performance score, and the composite score can be found in Figures 2 through 10c.

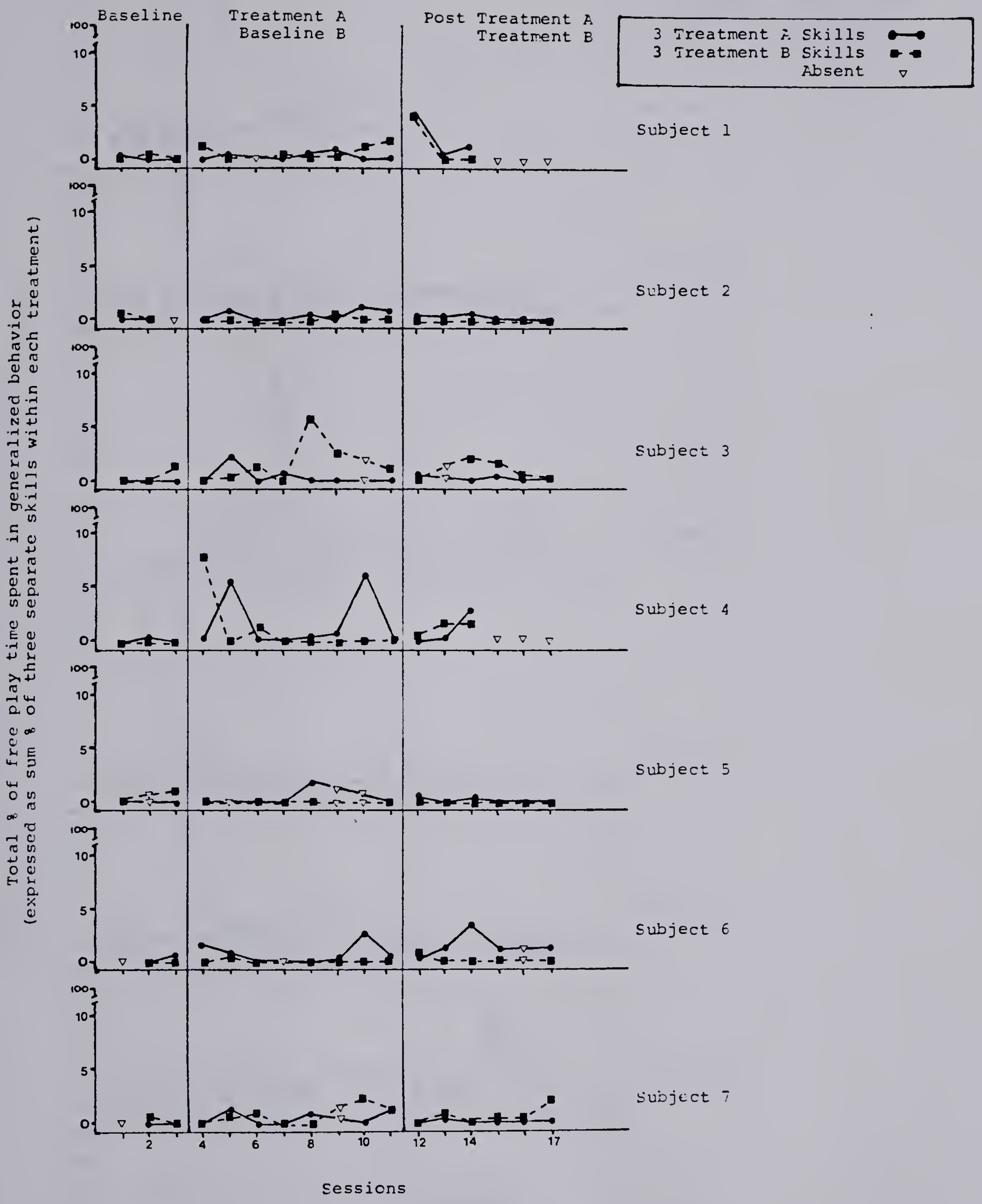


Figure 2 : Performance Graphs for the Category : Touch.

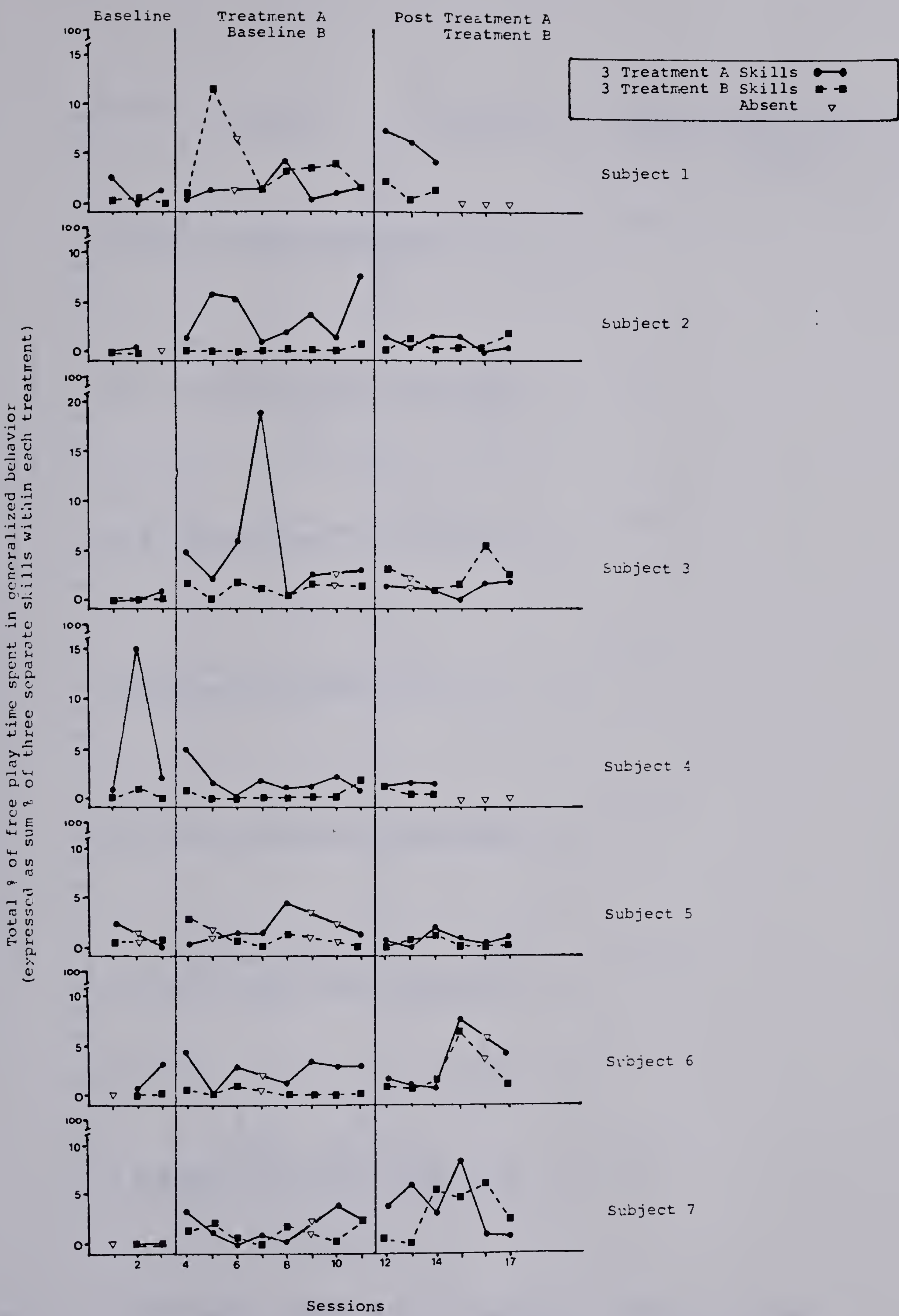


Figure 3 : Performance Graphs for the Category : Attend.

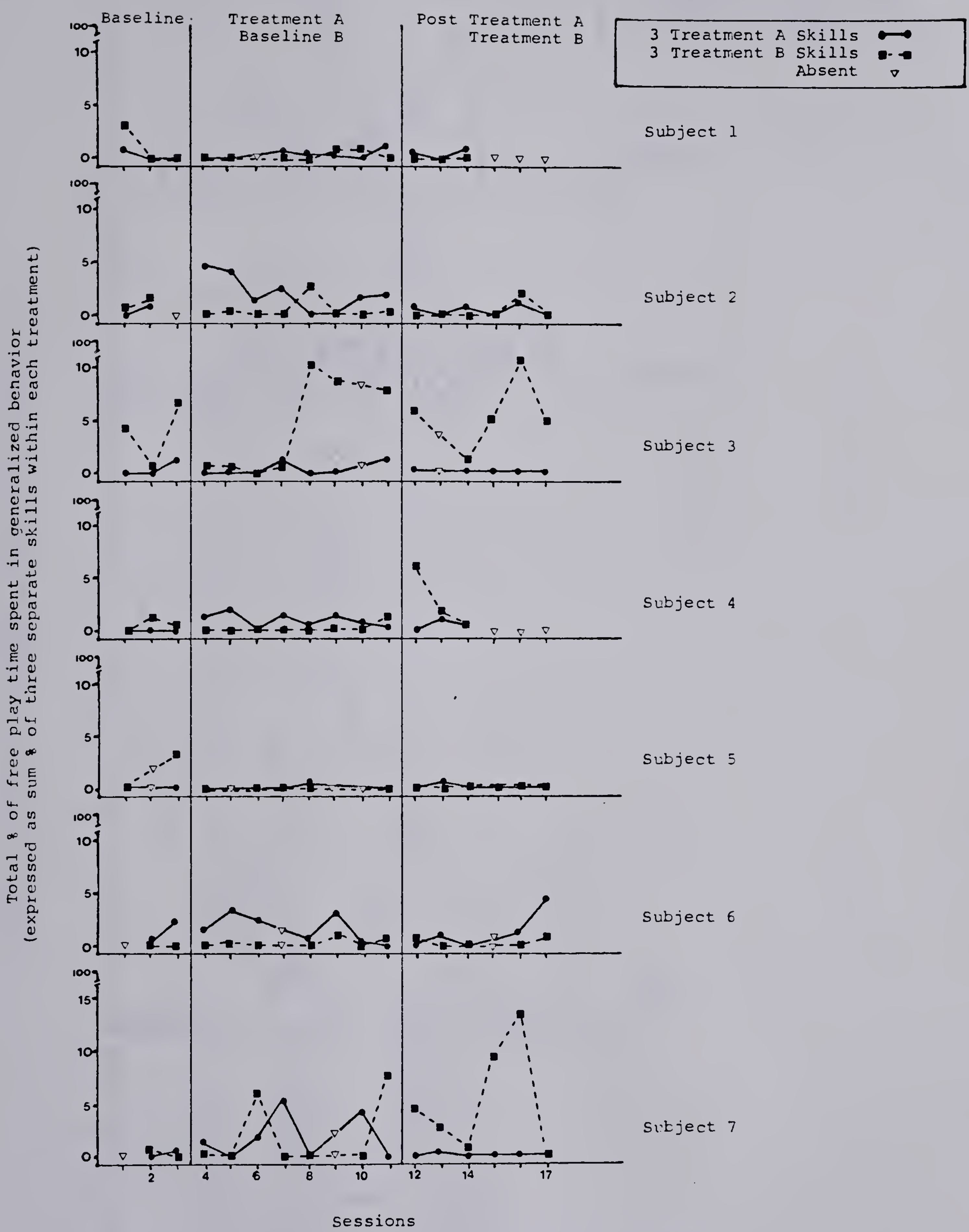


Figure 4 : Performance Graphs for the Category : Physical Contact.

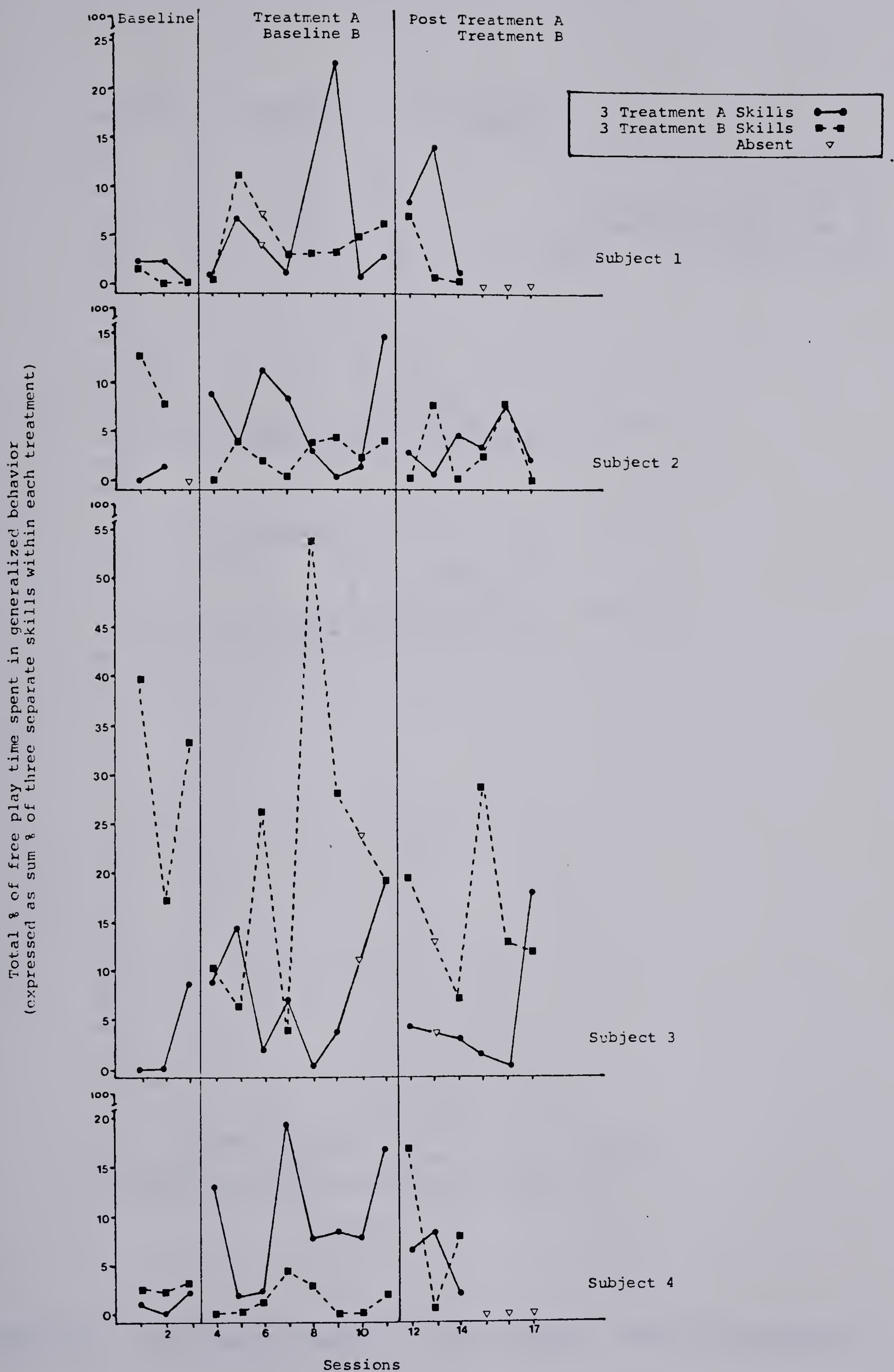


Figure 5a : Performance Graphs for the Category : Object Manipulation.

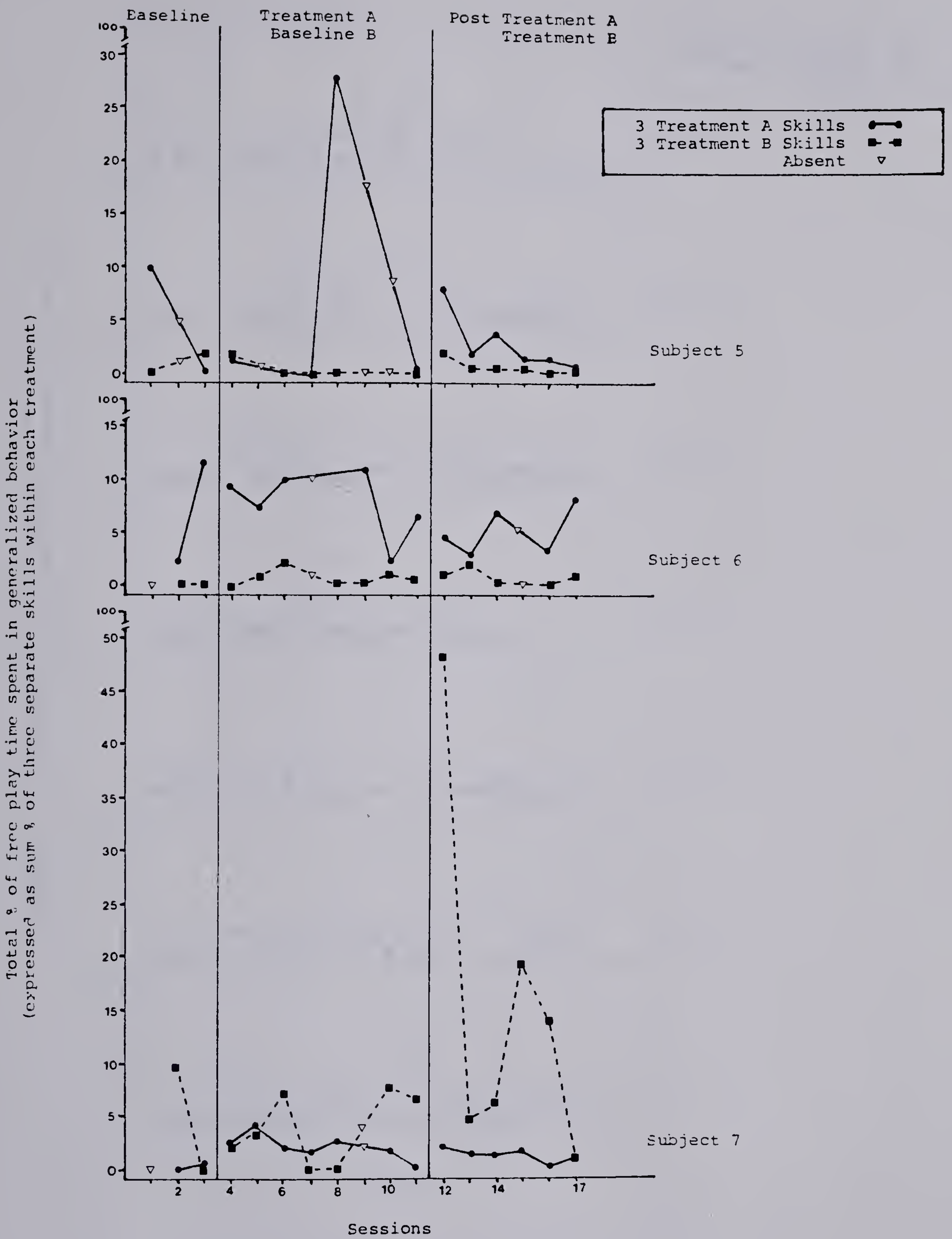


Figure 5b : Performance Graphs for the Category : Object Manipulation.

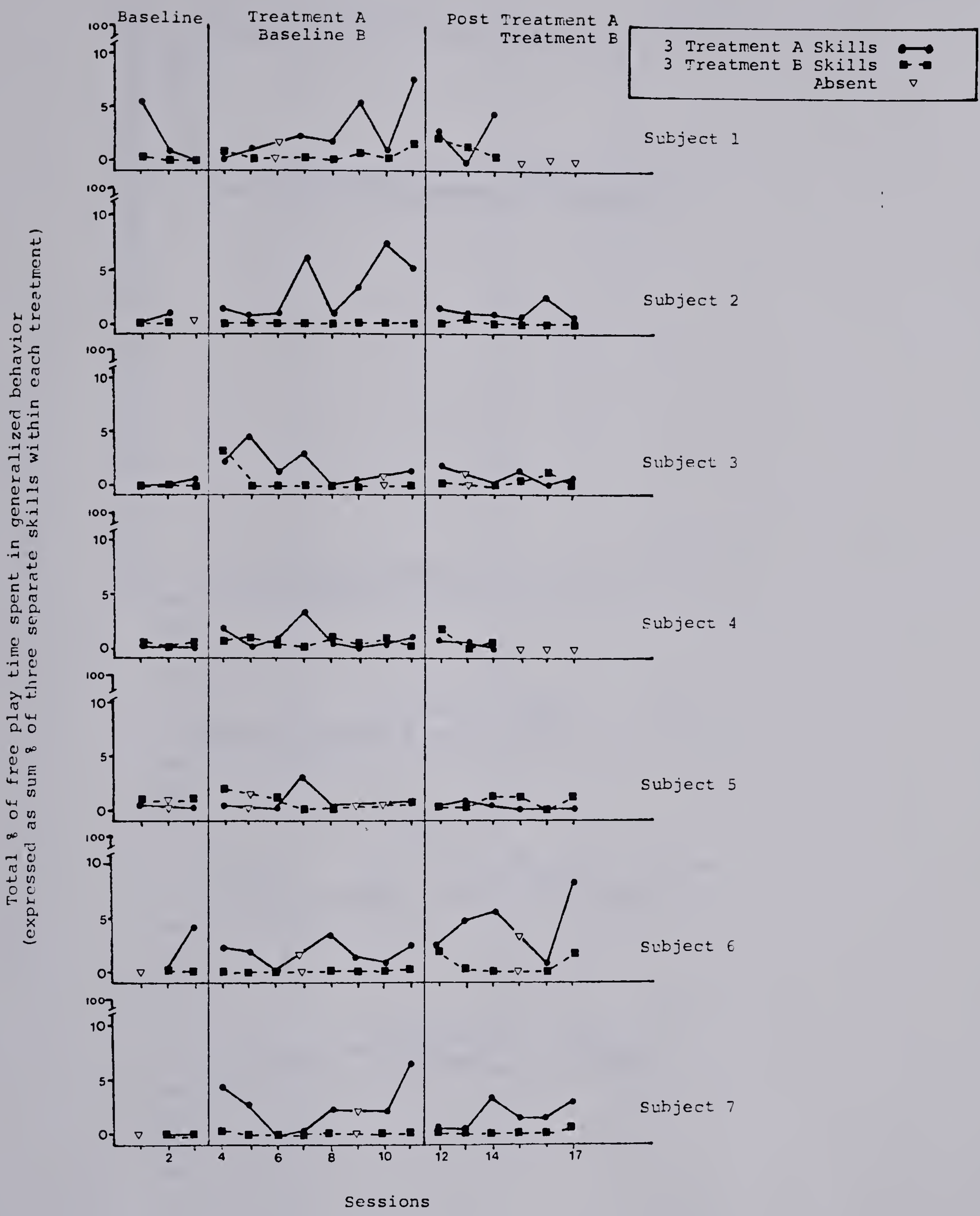


Figure 6 : Performance Graphs for the Category : Performance Below Last Level of Instruction.

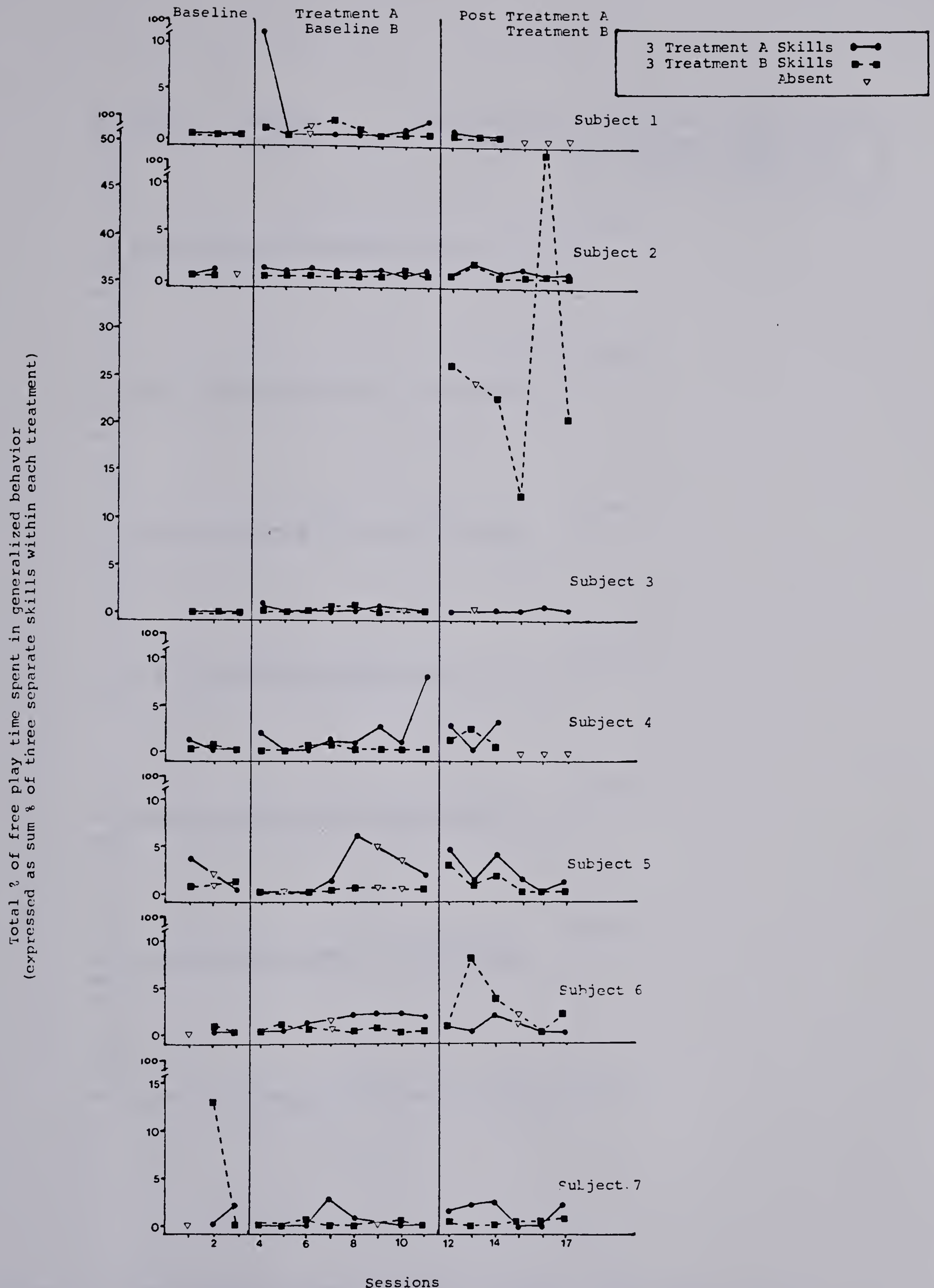


Figure 7 : Performance Graphs for the Category : Performance At Last Level of Instruction.

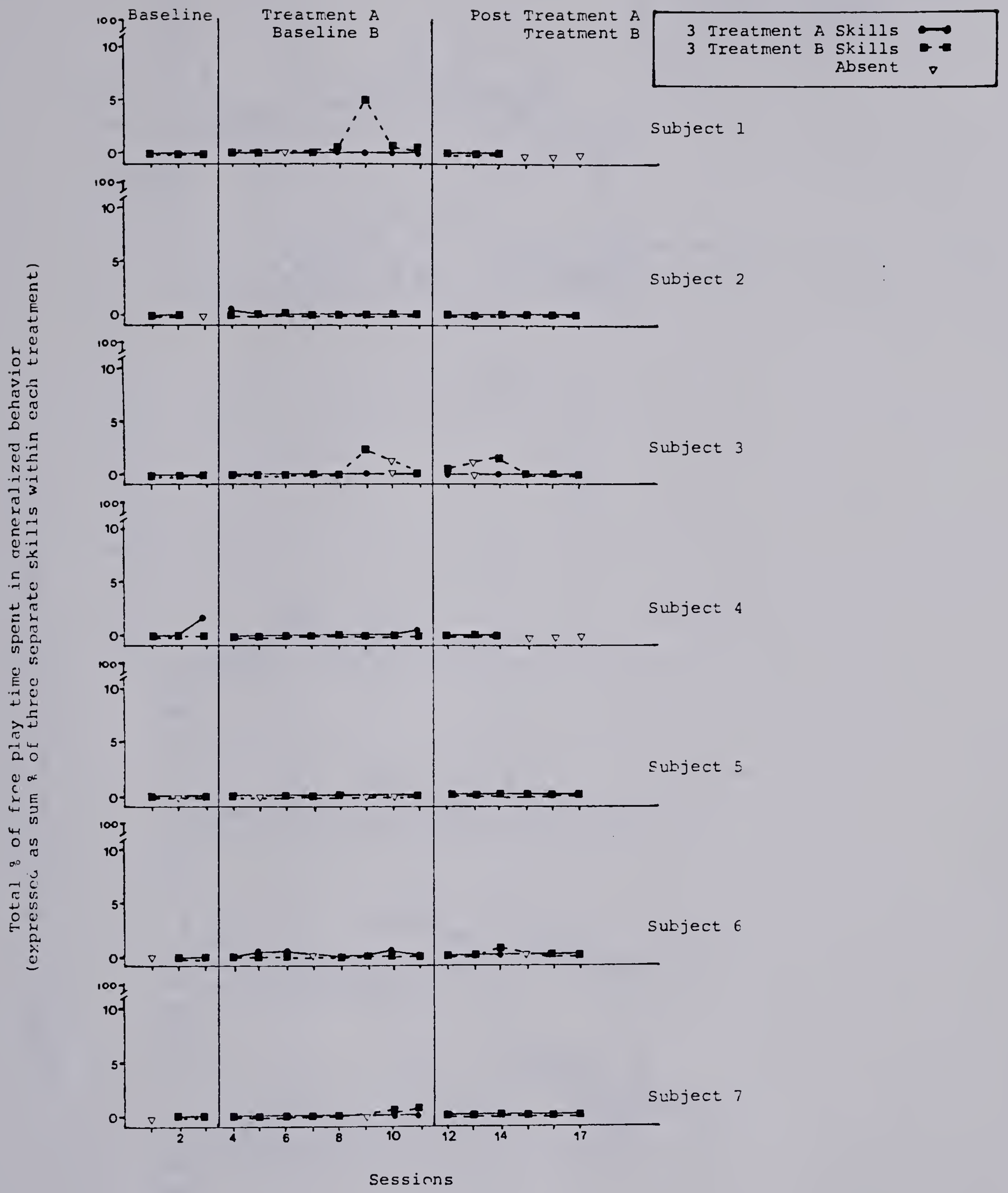


Figure 8 : Performance Graphs for the Category : Performance Above Last Level of Instruction.

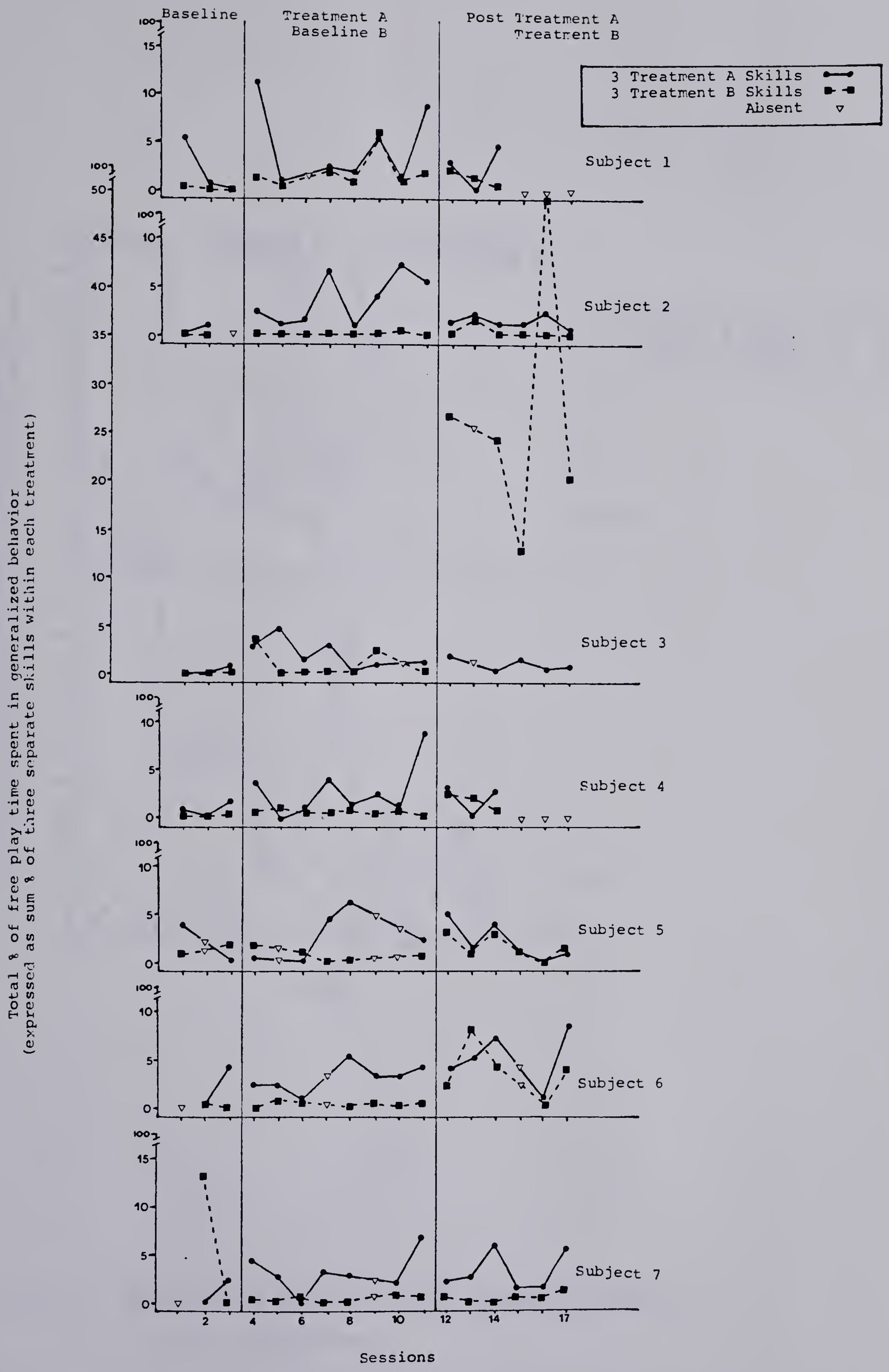


Figure 9: Performance Graphs for Skilled Performance Score: A Composite of Performance Below, At, and Above Last Level of Instruction.

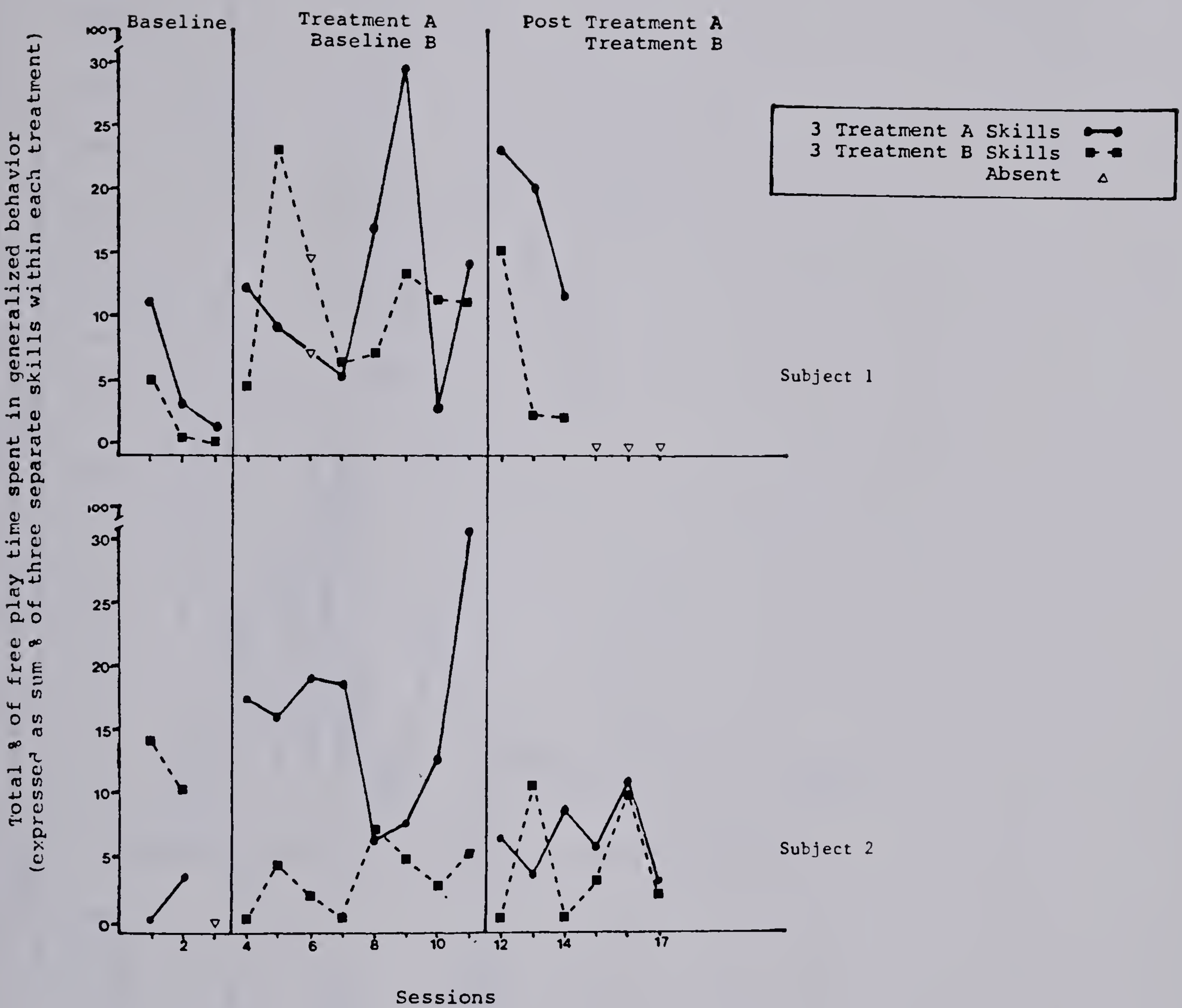


Figure 10a : Performance Graphs for the Composite Score of All 7 Categories.

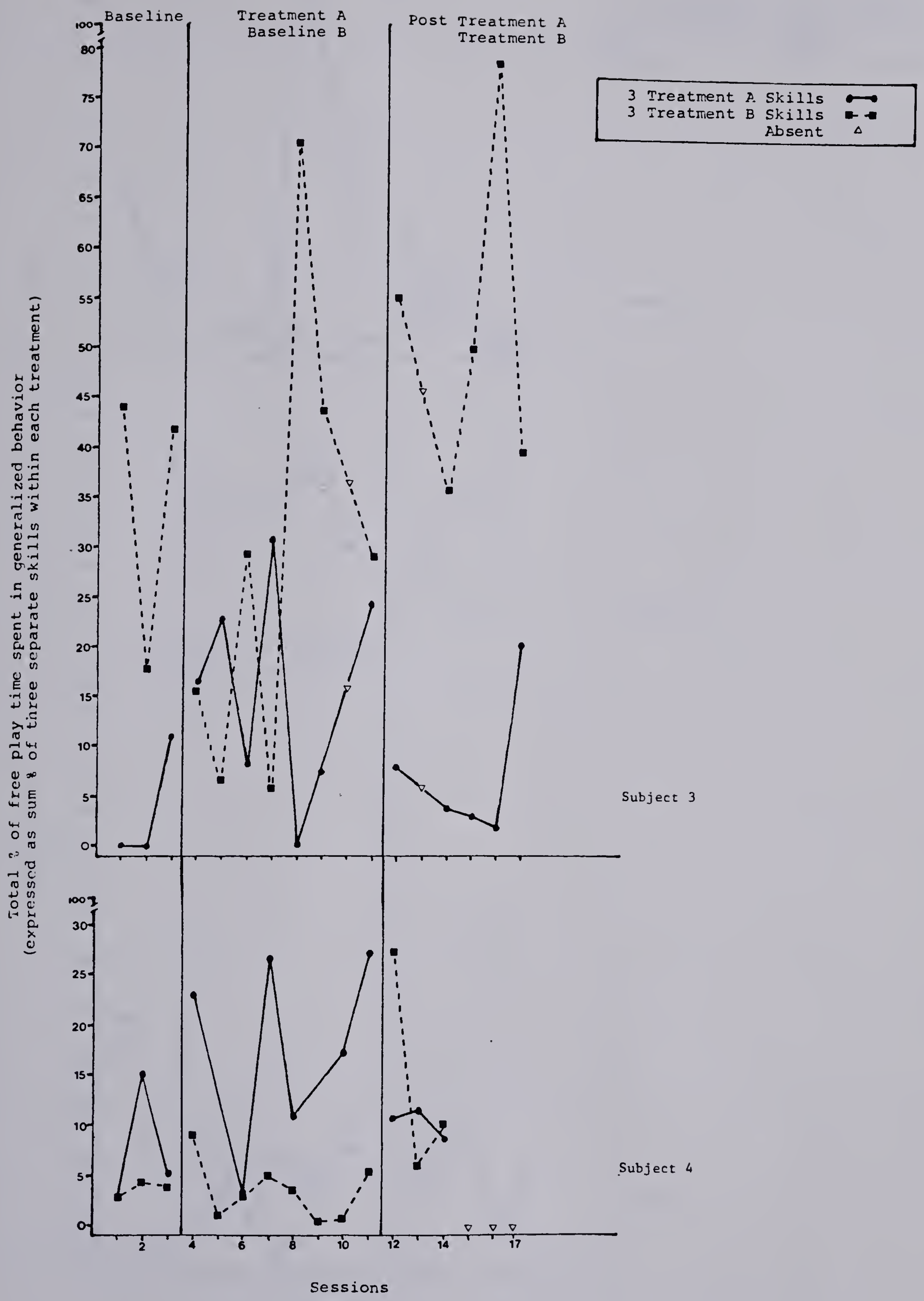


Figure 10b: Performance Graphs for the Composite Score of All 7 Categories.

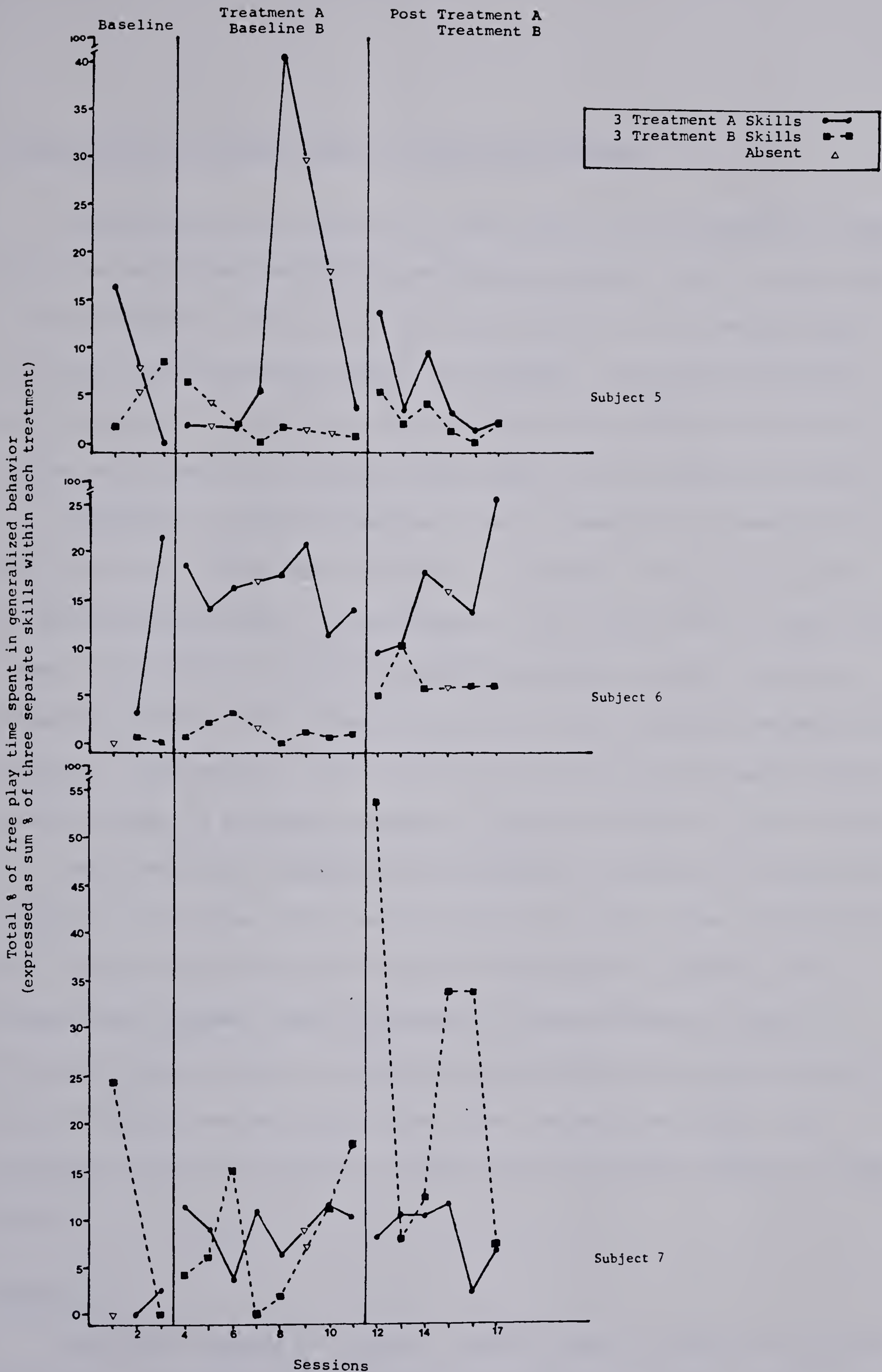


Figure 10c : Performance Graphs for the Composite Score of All 7 Categories.

Generalization of PREP Skills in Individual Subjects

Generalization was measured in this study as the percentage of free play time scored in the skilled performance category. That is, any time a child performed a skill at a task step below, at, or above the task step last receiving instruction or maintenance, the behavior was measured as generalized behavior. The generalization of PREP skills was analyzed on the basis of individual performance in six prescribed skills.

Performance records during the seventeen sessions of investigation are displayed individually in Figures 11 through 17. The vertical axis of each graph represents the percentage of free play time that the child chose to initiate at a level of skilled performance, each of the six prescribed skills. Units along the horizontal axis represent separate PREP sessions. This measure of free play was extracted from continuous observation records of 42 minute segments of activity within the PREP Program.

Each data point represents a proportion of the number of intervals in which a prescribed skill was initiated out of the number of intervals the child was observed free of teacher interaction on that day. Data within each treatment phase represent the generalization of skills in free play. Data in the post treatment phase indicate the generalization of skills which was maintained even after treatment was ended. Post treatment data were collected for those skills that were taught in Treatment A.

Subject 1

Performance graphs for Subject 1 can be found in Figure 11. The target skills of hanging from the knees on a horizontal ladder, jumping over,

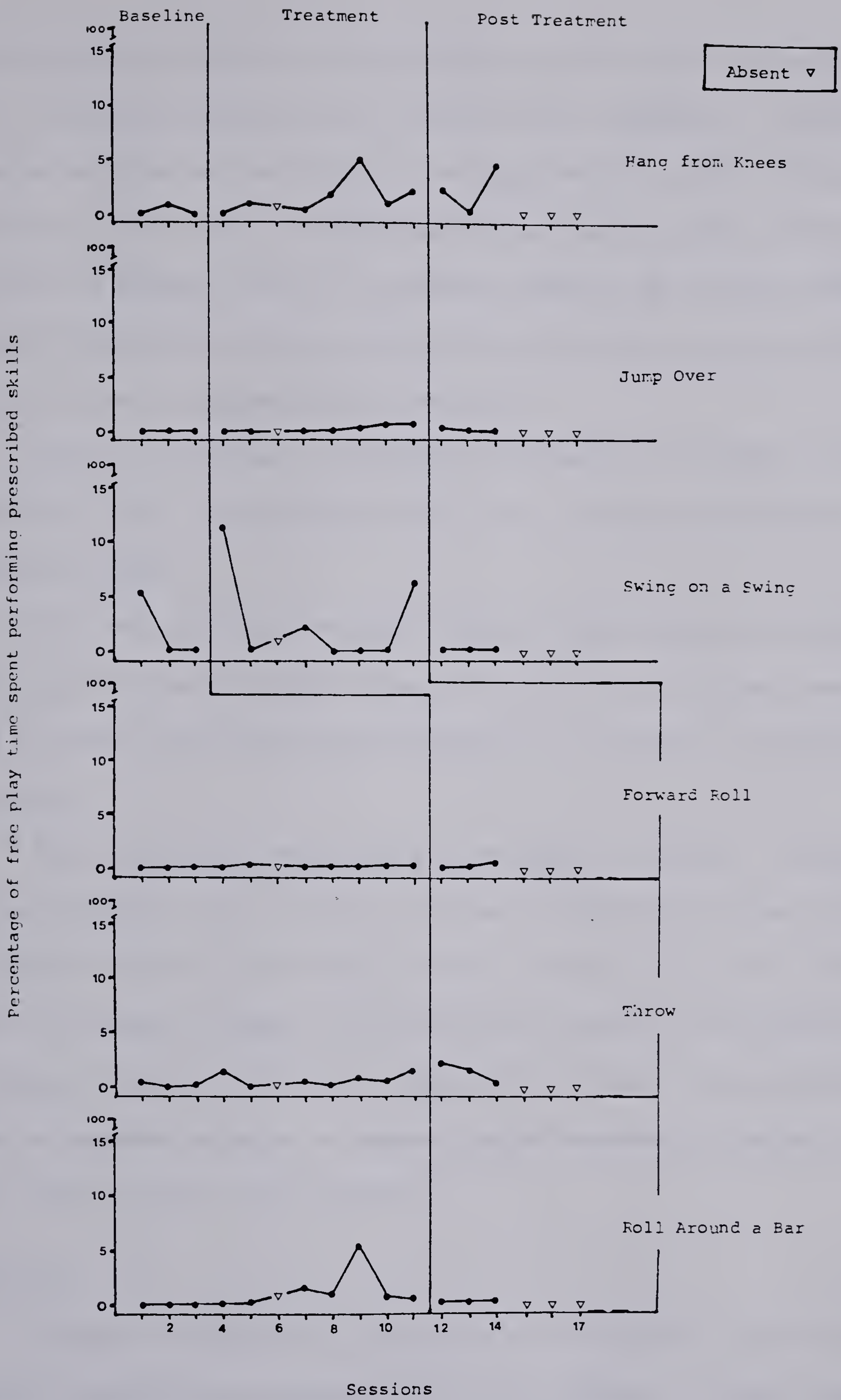


Figure 11 : Performance Graphs for Subject One.

and swinging on a swing were prescribed on the basis of pre-baseline free play inventory information and initial skill assessment. A measurable change was seen for the skill of hanging from the knees. The rate of initiation of the skill increased above baseline rate after Session 7 to a point in Session 8 where the subject spent 5% of his time hanging inverted. Performance remained above baseline rates in each successive session with the exception of Session 13.

Although there was an increased performance of jumping over during Sessions 10 and 11 (See Figure 11), it was too small to indicate strong positive change.

Data were variable for this subject's performance in swinging during baseline and treatment phases. His performance during Sessions 4 and 11 showed considerable generalization of the skill of swinging on a swing.

Three additional skills were prescribed for Subject 1 during the second treatment phase. Figure 11 clearly demonstrates that the skills of forward rolling and rolling around a bar were not used in free play while receiving treatment. While data for the skill of throwing showed a change in Session 12 when treatment began, there were insufficient data to support evidence of generalization because the subject was absent from the last three sessions.

Subject 2

As shown in Figure 12, generalization of treated skills was observed in one of six prescribed skills for Subject 2. During treatment of the first three skills, an increasing amount of time was spent sliding down the slide to a maximum of 7.1% in Session 10. Post treatment

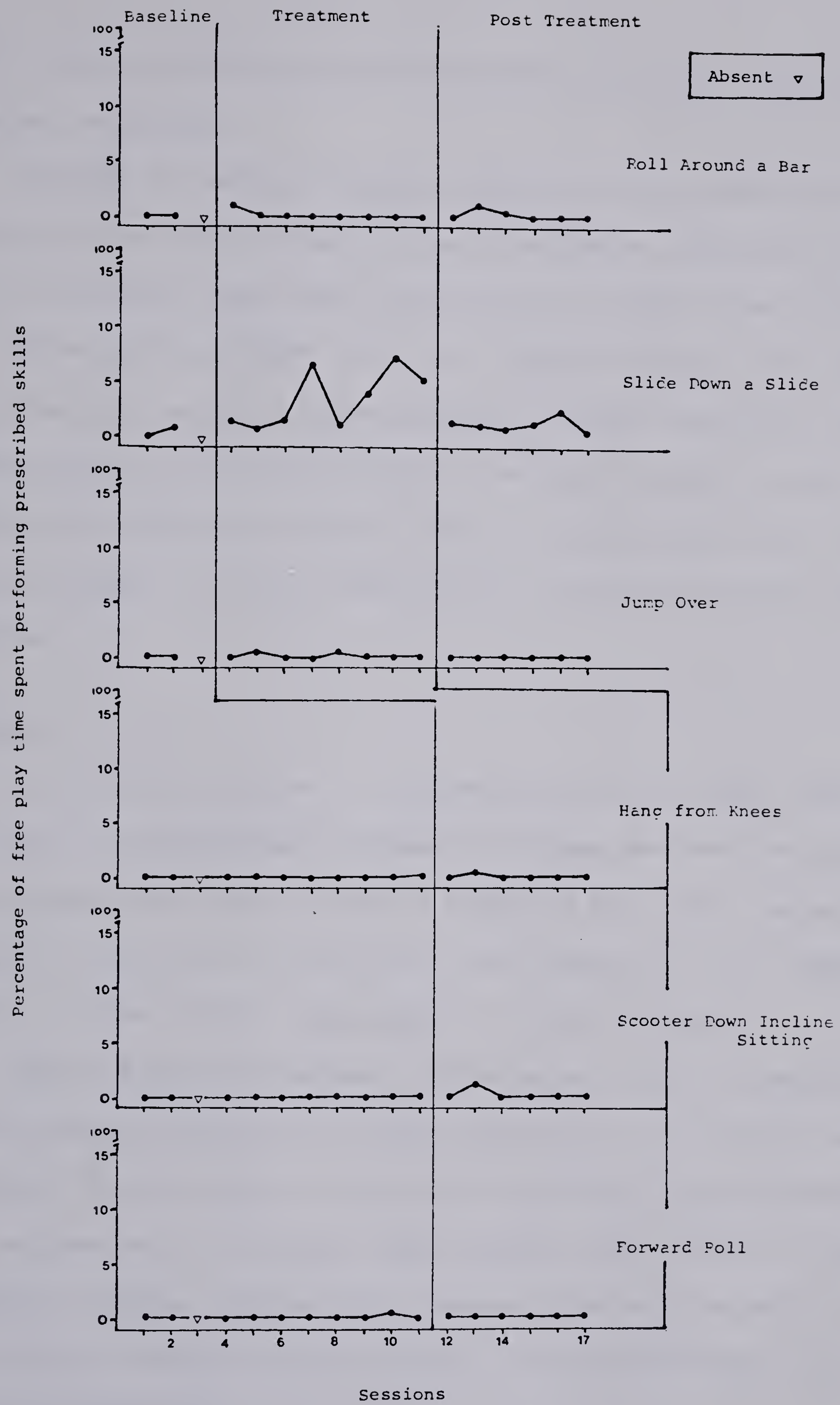


Figure 12 : Performance Graphs for Subject Two.

data in every session demonstrated a positive difference from the average baseline performance.

Although the skills of rolling around a bar and jumping over were initiated on occasion during treatment, performance frequency and duration of these two skills were insufficient to suggest generalization.

Three additional PREP skills were treated during a second treatment phase however, Subject 2 seldom chose to initiate these skills. No generalization of the target skills of hanging inverted, scooter riding, or forward rolling was observed. During seventeen days of data collection, performance of each of these skills was observed on only one occasion.

Subject 3

A graphic illustration of the generalization of skills prescribed for Subject 3 is presented in Figure 13. During the first two sessions of treatment, this subject showed a change in time spent jumping on the trampoline from 0% of his free play time to nearly 2%. This change indicated a change in free play behavior, at least initially.

Subject 3 received treatment in the target skill of jumping over. He was observed initiating the skill during only one of seven sessions attended. The percentage of time spent in sliding activity during free play was measured to be greater than baseline time in five of seven treatment sessions. Although post treatment data were variable, an apparent change between baseline and post treatment performance was indicated in Figure 13.

Subject 3 clearly demonstrated a change in the initiation of scooter riding during treatment. Generalization of this skill was pronounced.

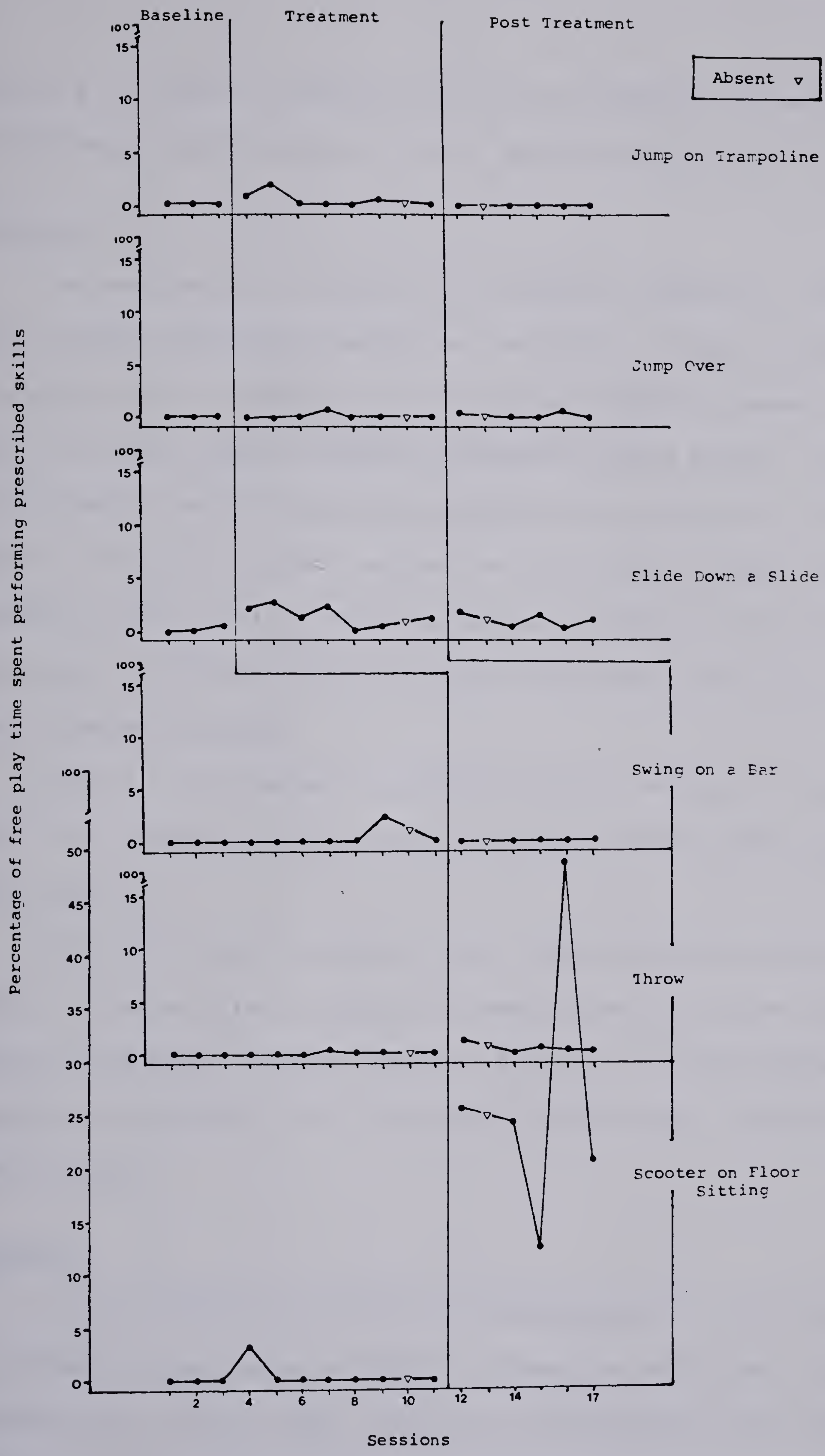


Figure 13 : Performance Graphs for Subject Three.

Data for the skills of swinging on a bar and throwing were equally as convincing in demonstrating a lack of generalization.

Subject 4

The performance of Subject 4 is recorded in Figure 14. The first three skills receiving attention were ascending a ladder, jumping on a trampoline, and swinging on a bar. Baseline variability makes it difficult to compare ladder climbing performance between phases. Time spent in initiated trampoline jumping increased with treatment to 7.5% in Session 11. Percentage scores improved over the stable baseline zero performance in 75% of all treatment sessions. In addition, this skill was initiated in 1.2% and 2.5% of free play time during two of the three post treatment sessions.

Subject 4 was observed initiating swinging activity on a bar during half the treatment sessions. The percentage scores for these days were all below 1%.

During the second treatment phase, three new skills were prescribed. Figure 14 demonstrates a change for jumping down and scooter riding as intervention began. Because Subject 4 dropped out of the program two weeks early, available data are insufficient to permit comparison between phases.

Subject 5

Generalization of two of the six skills prescribed for Subject 5 is evident in the graphs in Figure 15. These two skills were ladder climbing and scooter riding. Although the percentage of time spent climbing a ladder did not increase immediately after treatment began,

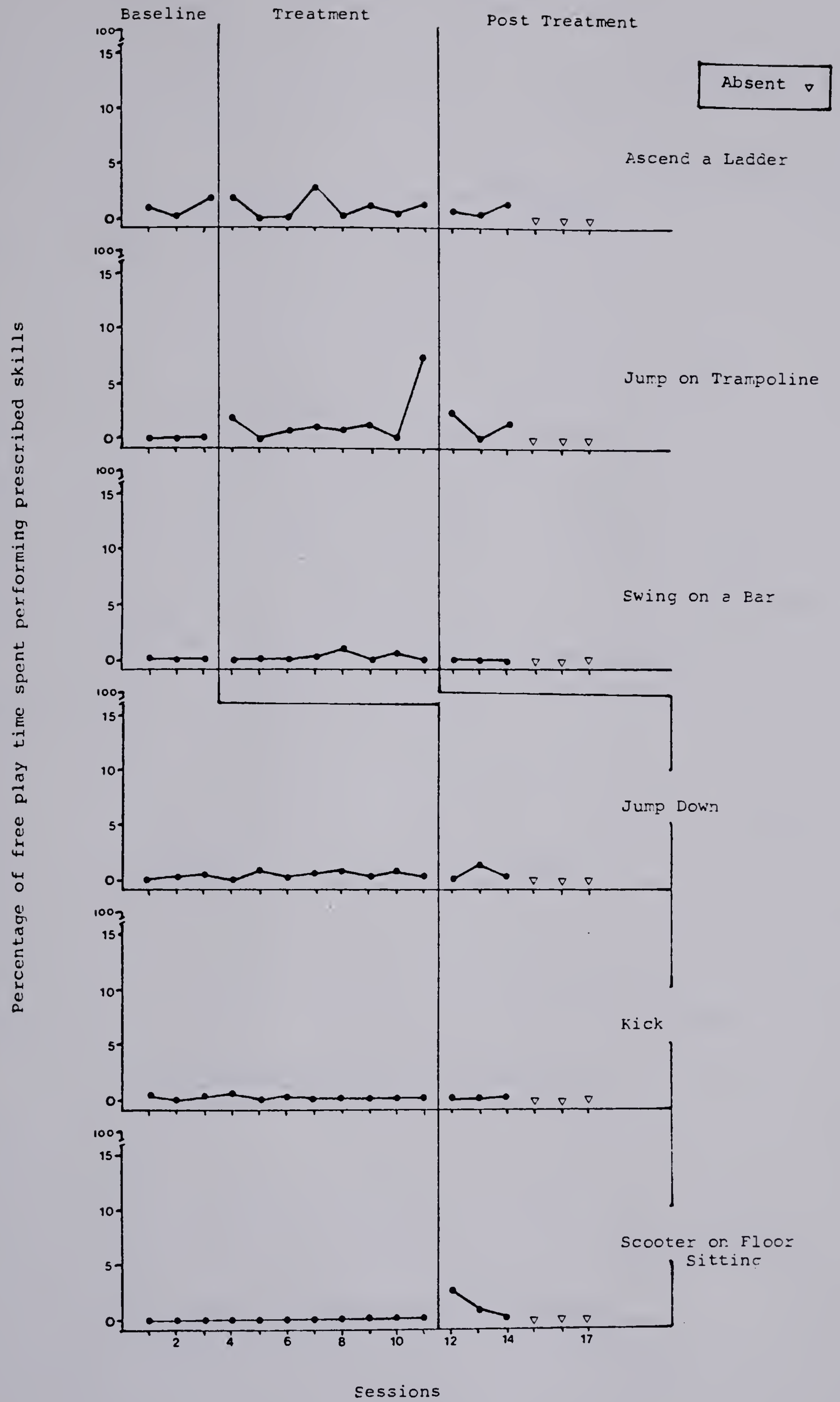


Figure 14 : Performance Graphs for Subject Four.

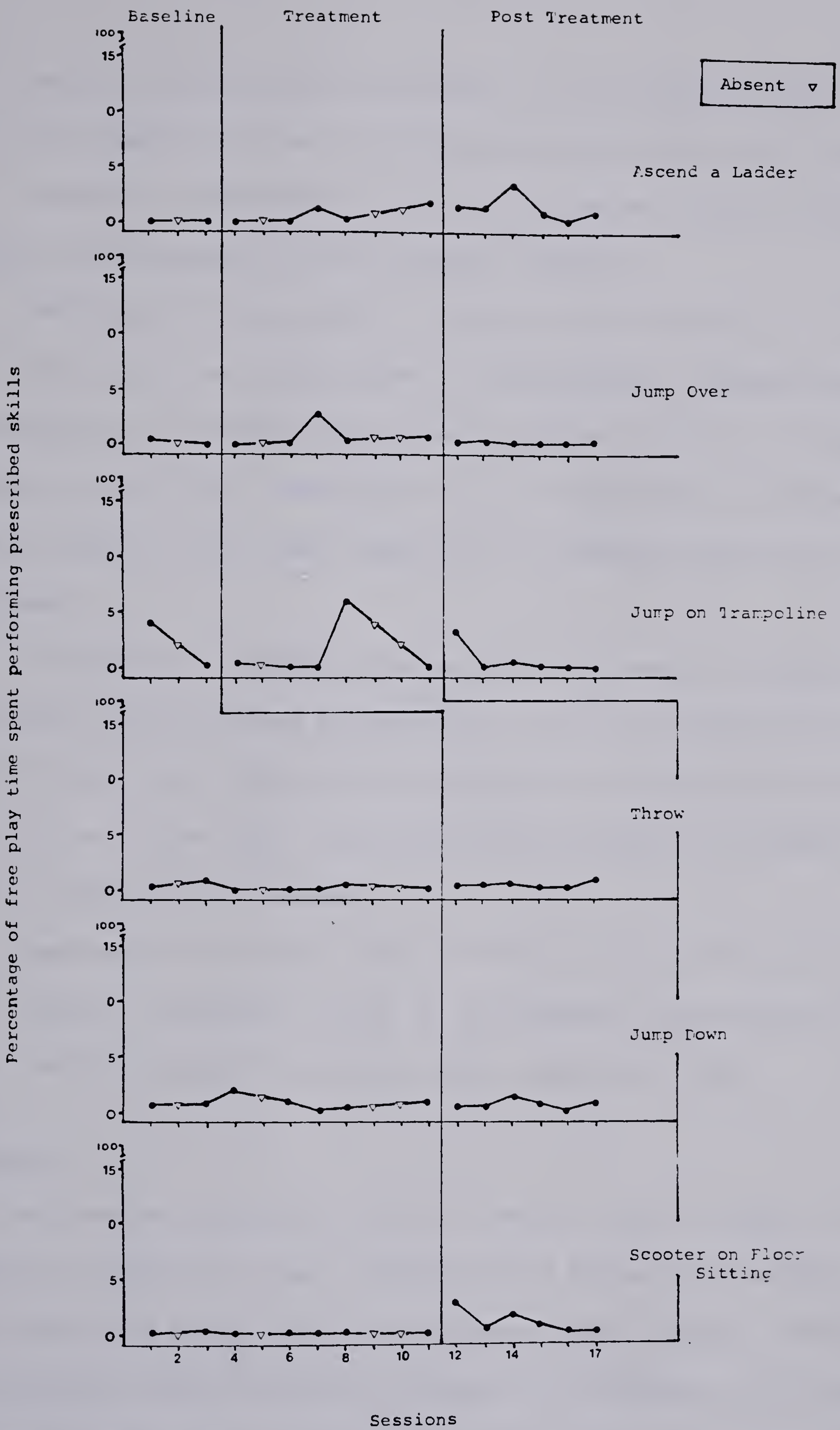


Figure 15 : Performance Graphs for Subject Five.

performance during treatment sessions 7, 8, and 11 indicated a change from zero baseline scores. Performance was maintained after treatment was terminated in Sessions 12, 13, and 14, demonstrating a definite change from baseline to post treatment behavior.

Performance of the skill of jumping over in Session 7, in combination with the recorded initiation in the remaining treatment sessions, demonstrated an increase above baseline scores. However, the over-all change was not large. Variability in all phases makes it difficult to assess change in time spent jumping on a trampoline during and after treatment.

The skills of throwing, jumping down, and scooter riding received treatment during the last six sessions. As shown in Figure 15, Subject 5 seldom threw a ball while playing freely even though she was receiving instruction in the skill. Similarly, her performance in jumping down was not affected by treatment.

Observation of scooter riding initiation in two thirds of the treatment sessions indicated a change in performance from baseline to treatment. Baseline values of zero make this change very clear.

Subject 6

Performance graphs for Subject 6 can be found in Figure 16. The skills of swinging on a bar, sliding down a slide, and jumping over were prescribed during the first treatment phase. Changes were observed on all three skills. As shown in Figure 16, initiation of bar swinging improved greatly during Sessions 10 and 11. Similar generalization of this skill was observed only once during the post treatment phase.

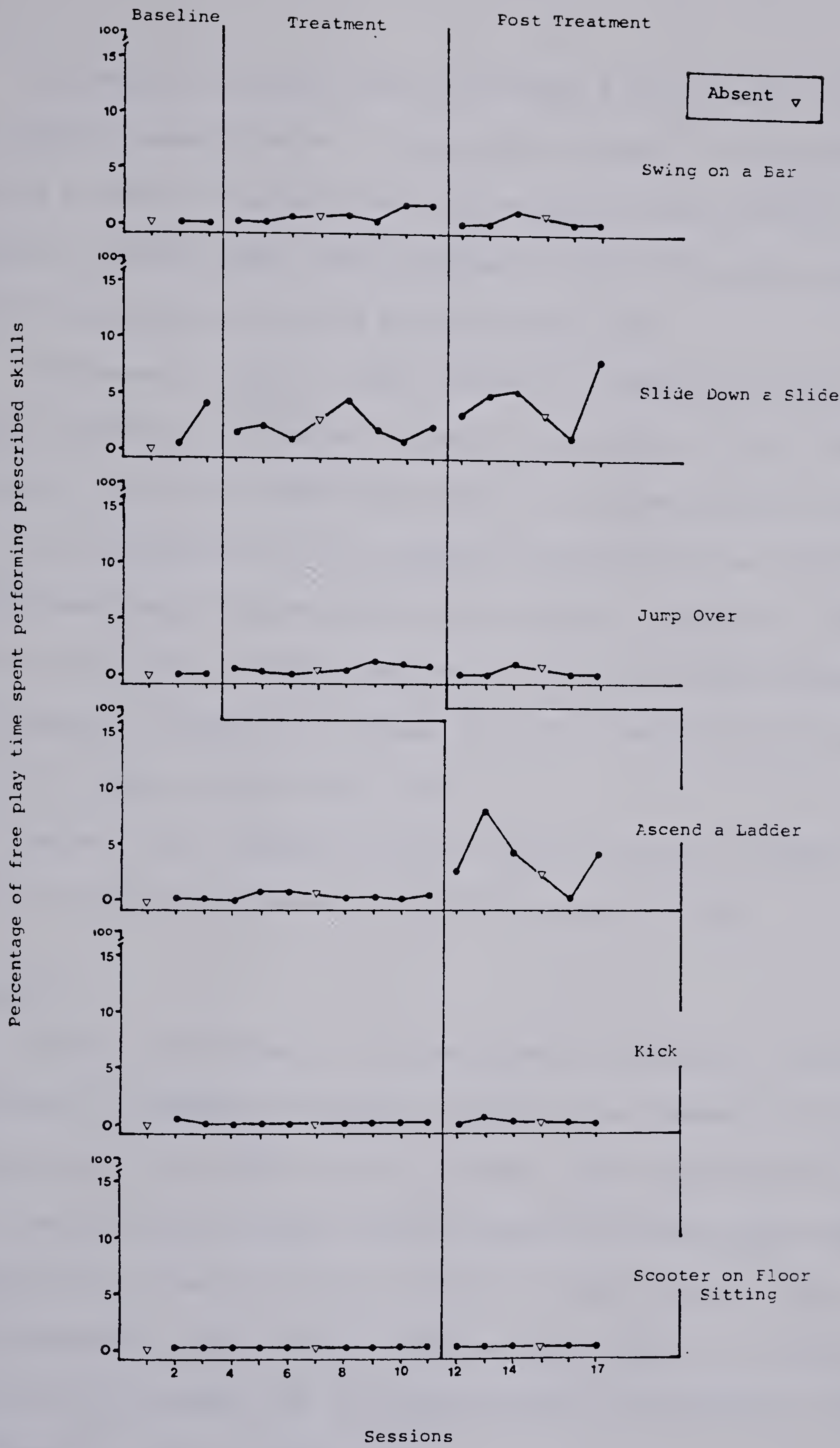


Figure 16 : Performance Graphs for Subject Six.

Although the baseline data for Subject 6 are variable for the skill of sliding, generalization is suggested in Figure 16. Initiation rate during treatment remained above zero percent in every session and continued to increase even after treatment. Four of five post treatment scores were greater than the mean baseline score.

Performance scores for this subject in jumping over were greater during treatment than before treatment. The extent of this change was limited, as the percentage score rose above 1% during only one session.

Of the three new skills treated during the last six sessions, a significant change was observed for one skill, ascending a ladder. Initiation of ladder climbing increased during the first treatment session and was recorded as a minimum of 2.5% of Subject 6's free play time in every session except Session 16.

Subject 6 was observed kicking only once during treatment and she was not observed initiating the skill of scooter riding.

<u>Subject 7</u>

Graphic illustration of the performance of Subject 7 is presented in Figure 17. Evidence of change in behavior was demonstrated for two of the first three skills treated. Clearly, the percentage of time spent walking up an incline during treatment increased above the zero baseline rate. Data in the post treatment phase show that this change was maintained. A more subtle change was indicated in the throwing graph, with increases over the mean baseline score evident in both treatment and post treatment phases.

In addition, generalization of ladder climbing was suggested. Close examination of the graph in Figure 17 revealed that treatment data ex-

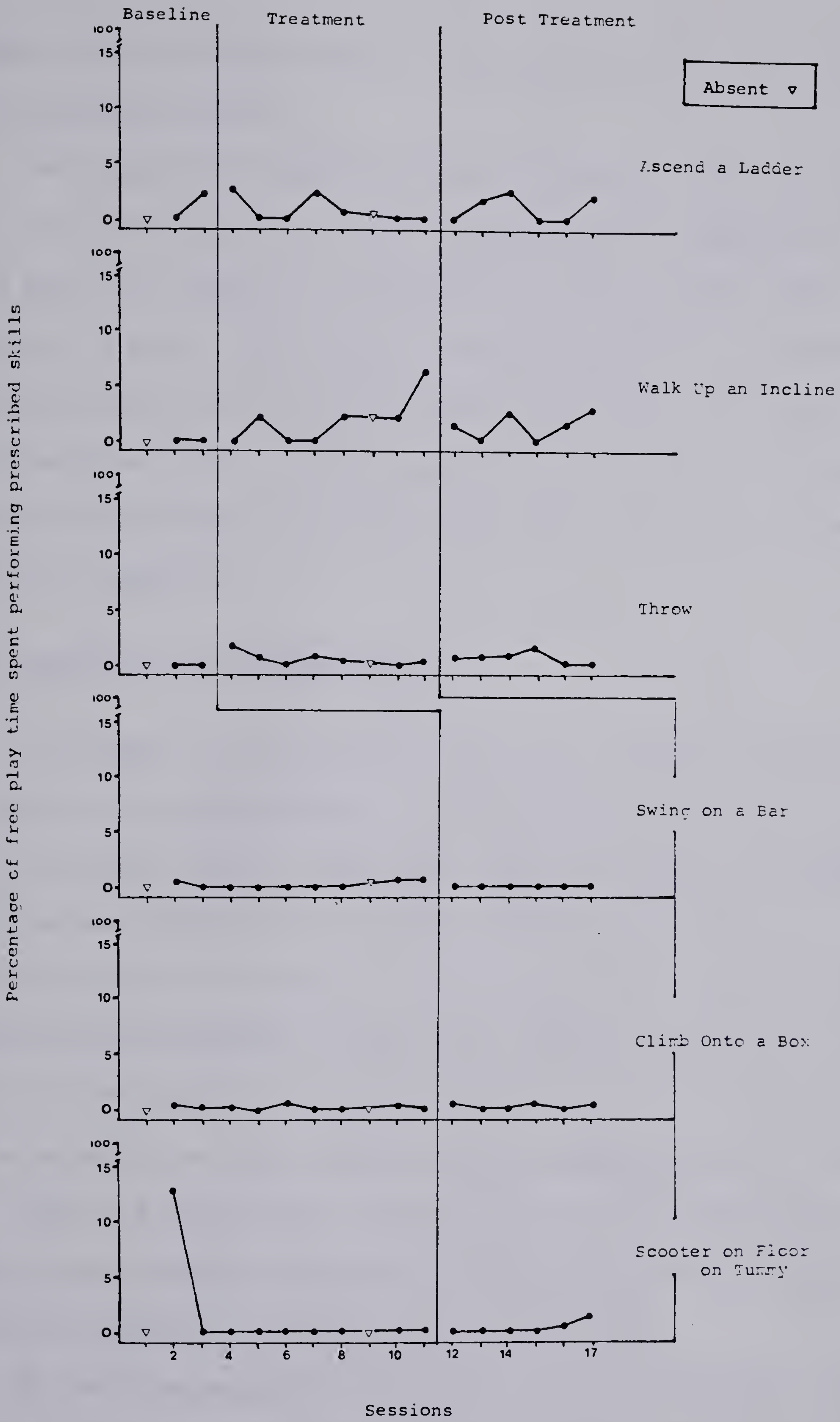


Figure 17 : Performance Graphs for Subject Seven.

ceeded the mean baseline score in two treatment sessions and three post treatment sessions.

The bottom three graphs in Figure 17 indicate that generalization did not occur for any one of the prescribed skills during the last treatment phase. The subject was observed initiating scooter riding during Sessions 16 and 17, after four treatment sessions. On the basis of the available data it may be concluded that treatment of the three skills of swinging on a bar, climbing onto a box, and riding a scooter did not significantly change the amount of free play time spent in these activities for Subject 7.

Skill Specificity of Generalization

In looking at the changes in time spent initiating treated skills, it appeared that generalization may be skill specific. In order to examine this more closely, performance graphs for skills prescribed to three or more children were grouped together. This information, provided in Figures 18 through 24, allowed analysis of specific skill generalization for the group. Generalized behavior was measured as percentage of free play time spent initiating the skill at a level of performance below, at, or above the last level of instruction for each day.

Figure 18 illustrated a change in skilled performance with treatment for every subject prescribed the skill of ascending a ladder. The amount of change was variable across individuals.

Of the three subjects prescribed the skill of jumping on the trampoline, Subject 4 was the only one demonstrating a positive change from baseline performance.

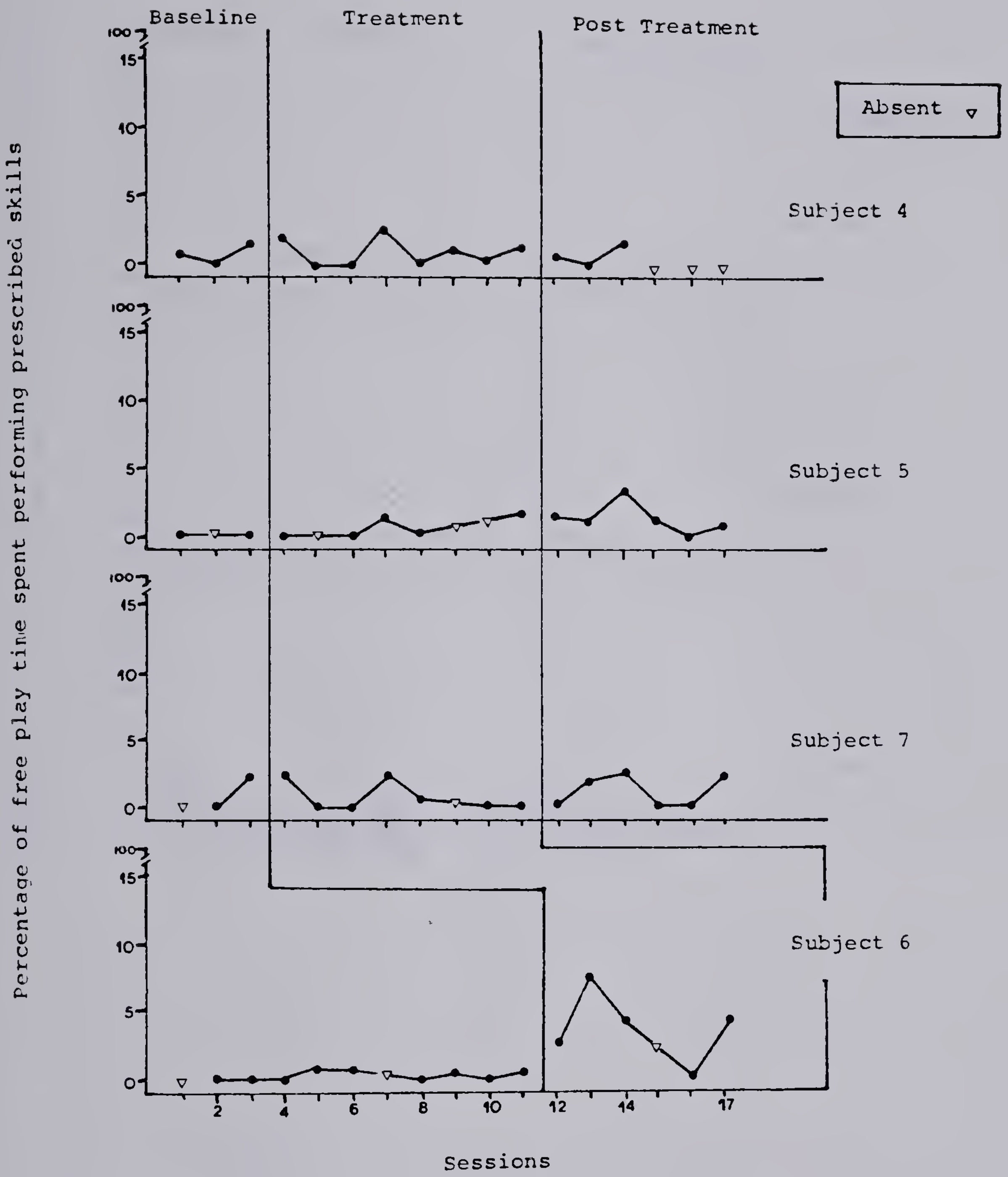


Figure 18 : Performance Graphs for all subjects prescribed the skill : Ascend a Ladder.

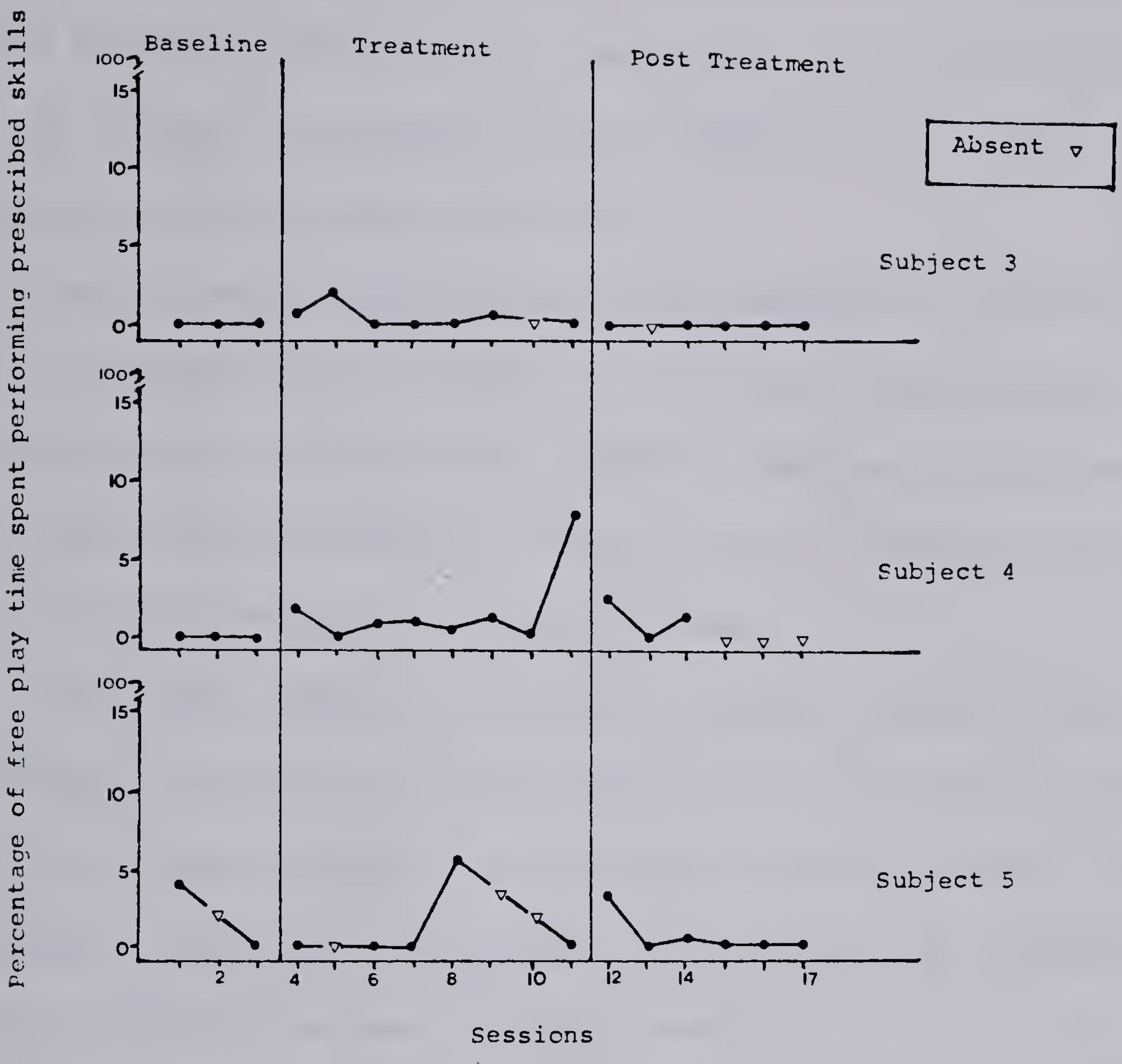


Figure 19 : Performance Graphs for all subjects prescribed the skill : Jump on the Trampoline.

Performance graphs in Figure 20 indicate that while the skill of jumping over was seldom initiated during baseline, initiation was observed for each subject during treatment, and on more than one occasion in the post treatment phase for Subjects 3, 5, and 6. However, percentage scores seldom reached 1%.

Clear change in performance with treatment is indicated for Subject 3 in Figure 21. In addition, two of the remaining three children prescribed the skill of riding a scooter increased their use of the skill when treatment began. Subject 6 was not observed initiating this skill during baseline or treatment phases.

Performance graphs in Figure 22 indicate that all three children prescribed the skill of sliding down a slide performed the skill in free play more frequently once instruction in the skill began. This change is greatest for Subject 2. Subject 6 maintained the increased initiation of this skill in the post treatment phase.

Observation of increased initiation of bar swinging was recorded in two of four subjects. Of these two changes, Subject 4 initiated the skill on only two occasions. Subject 6 gradually initiated bar swinging more often as treatment continued, but returned to baseline performance after treatment.

Baseline variability for Subjects 1, 3, and 5 made it difficult to evaluate change for the skill of throwing. For these three subjects, performance during treatment was very similar to baseline performance. Subject 7 initiated throwing in free play during and after treatment with percentage scores of 1.5% or better in Sessions 4 and 15.

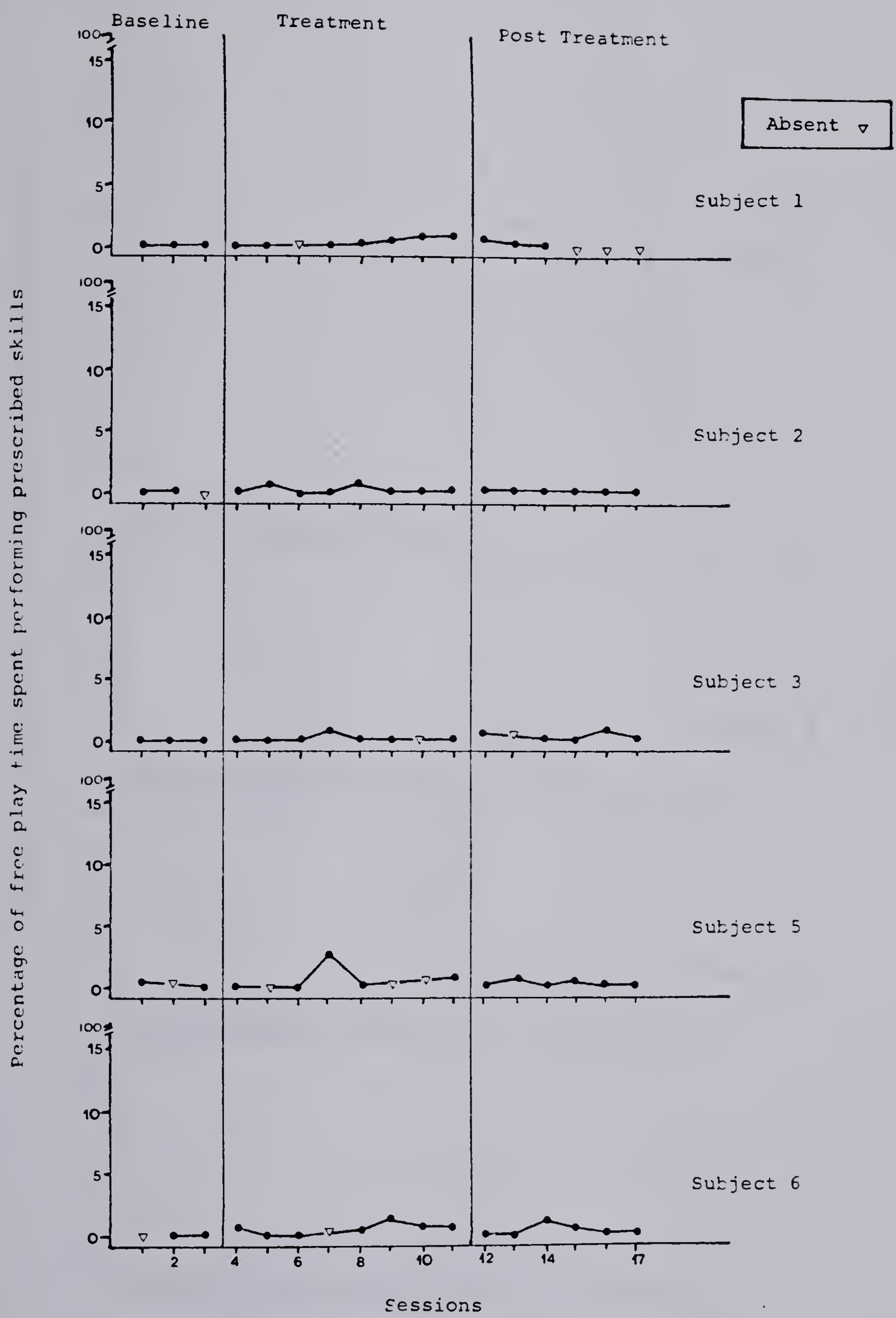


Figure 20: Performance Graphs for all subjects prescribed the skill : Jump Over.

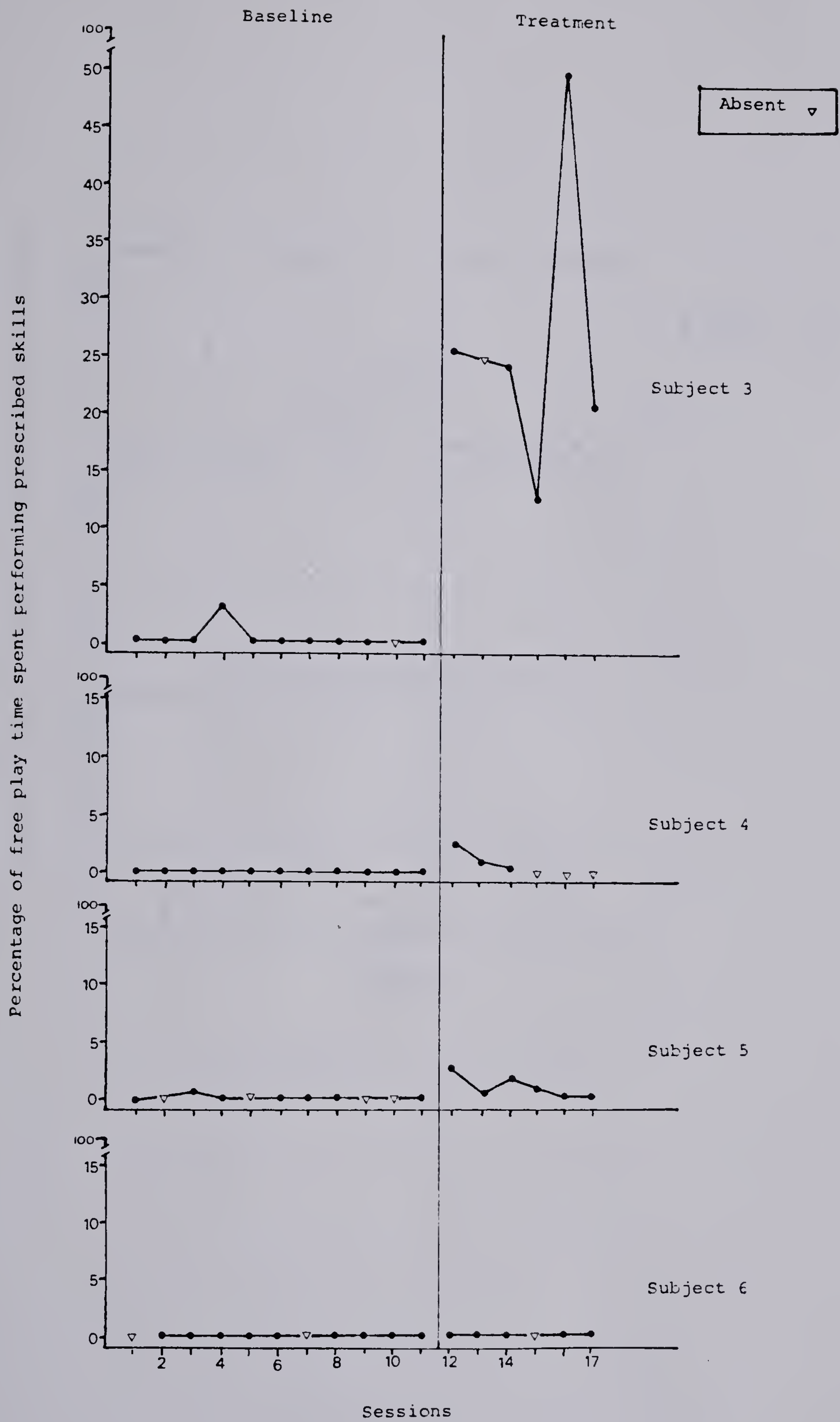


Figure 21 : Performance Graphs for all subjects prescribed the skill : Scooter Riding on the Floor on the Seat.

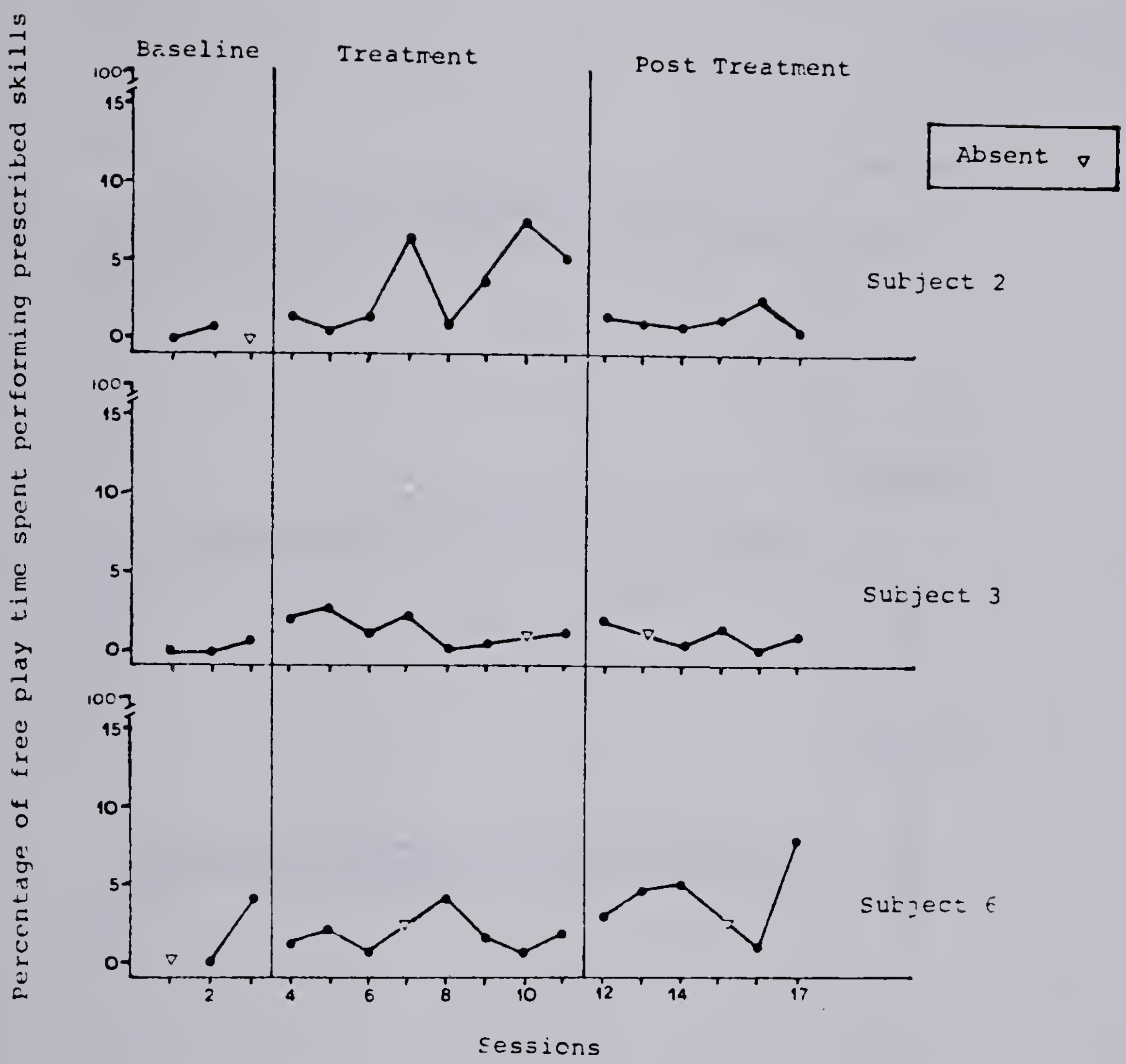


Figure 22 : Performance Graphs for all subjects prescribed the skill : Slide Down a Slide.

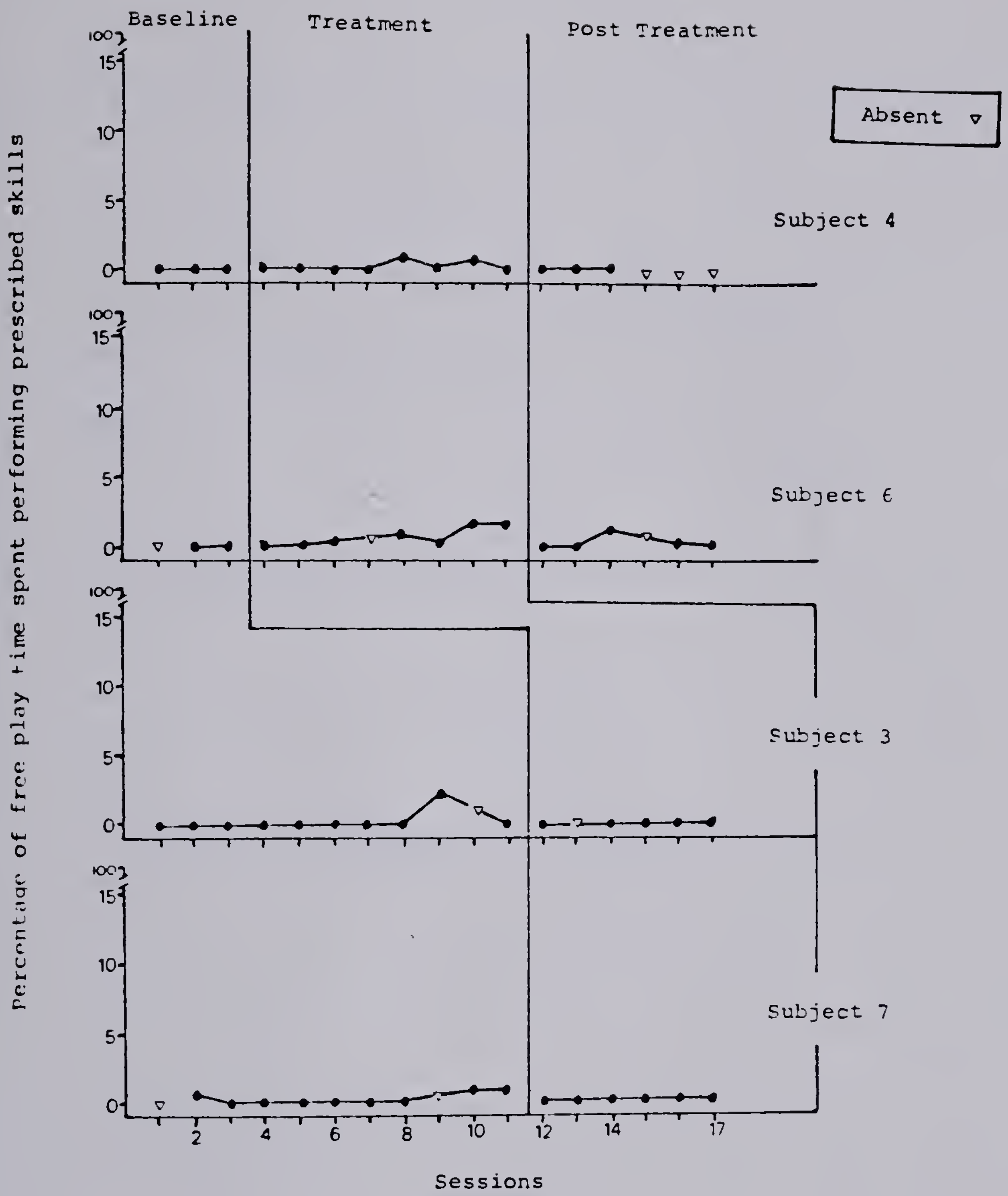


Figure 23 : Performance Graphs for all subjects prescribed the skill : Swing on a Bar.

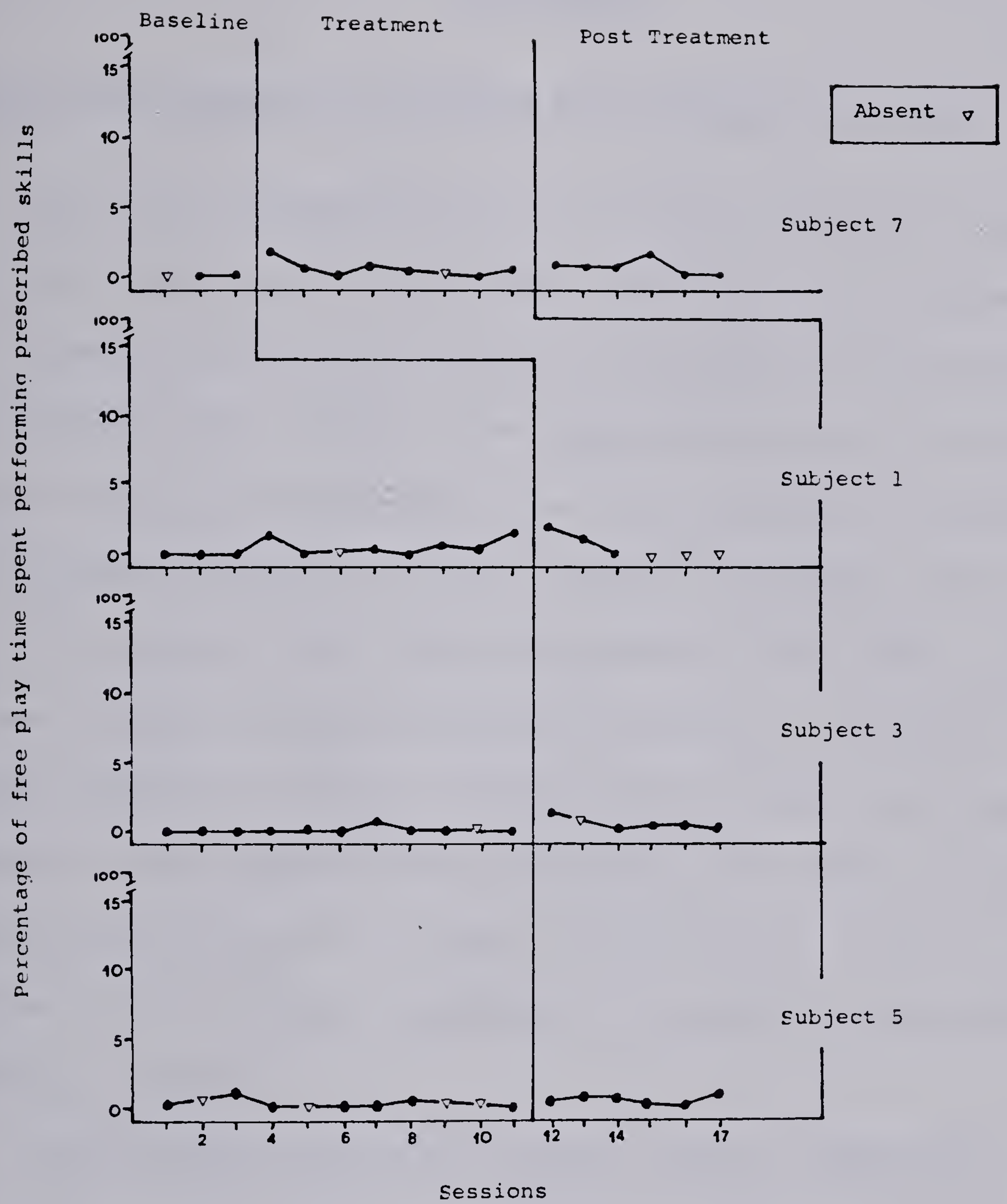


Figure 24 : Performance Graphs for all subjects prescribed the skill : Throwing.

CHAPTER V

DISCUSSION

Generalized Behavior as Described by a Category Instrument

The topic of generalization has received increased attention in recent years. Reasons for this could be many, but it is suggested that this attention is directly related to the growth in the applied behavior analysis field. Studies have repeatedly shown that stimulus generalization does not automatically occur as an outcome of a behavioral program (Jackson and Wallace, 1974; Rincouver and Koegel, 1975; Stokes, Baer, and Jackson, 1974; Walker and Buckley, 1972; Walker _et al_., 1975). Because stimulus generalization is so important to program success and because behaviorists were not meeting this program goal, they had to reevaluate their instructional strategies. As a result, methods of programming generalization have been developed in an attempt to improve the chances of seeing skilled performances in situations other than the experimental setting.

The behavioral literature provides limited information on the types of measures available to evaluate generalization. As a result, investigators have developed their own methods of evaluating the generalized training effects of their instructional programs. This may not be a problem for some researchers. For example, it may not be a problem in a program that sets out to reduce the head-banging behavior of a severely retarded child. After the behavior has been reduced or eliminated in the experimental setting, the child is observed in a variety of typical everyday situations. Two possible observations exist: either

the child reverts back to banging his head or he does not.

This is not always the case. In instructional programs such as PREP it is not so clear. When a child is being taught to jump on a trampoline and he is observed touching the trampoline or watching other children jump, is that a generalized training effect? Can performances which approximate but do not meet the specifications for a target performance be considered generalized behavior?

Because this problem has not been adequately addressed, a category observation instrument was designed to identify a list of behaviors which presumably described skilled performance and all possible approximations to the target performance. An attempt was made to include behaviors that have been previously used in the literature as evidence of generalization (Buell, Stoddard, Harris, and Baer, 1968; Flavell, 1973; Hall and Broden, 1967; Hardiman et al., 1975; Johnson et al., 1966). The instrument included eight behavior categories arranged in an order of sophistication from touching a piece of equipment on which the child was receiving instruction to actually performing the skill at a level above the last task step which received instruction.

Use of this detailed continuum of approximations to skilled generalization permitted the investigator to look at different possible dependent variables for use in describing generalized behavior. It was suspected that generalization of specific motor skills could be measured by a variety of active and passive behaviors. While it was felt that skilled performance, that is, actually performing the skill receiving instruction, would be a more realistic estimate of generalized behavior, collection of data across all behavior categories was carried

out in order to establish the relative usefulness of the behavior categories.

A pilot study (Terry, 1980) indicated that behaviors would be observed in all defined categories. Figures 2 through 10c show that the spectrum of behaviors related to the prescribed skills was observed, indicating that the instrument designed for this study was appropriate for describing the observed behaviors. It can only be assumed that these behaviors actually constituted generalization. However, this seemed to be a fairly safe assumption, since many of the categories described behaviors which have previously been used to measure generalization (Buell et al., 1968; Flavell, 1973; Hall and Broden, 1967; Hardiman et al., 1975; Jackson et al., 1966).

Expectation of generalization in PREP seemed realistic, especially because PREP teaching techniques incorporate generalization programming methods. Watkinson (1977) found that children in PREP spent an increased amount of time in skilled behaviors in her posttest assessment of free play, as well as an increase in the number of skills initiated. The data indicated that both stimulus generalization and response generalization seemed to occur in PREP. Consequently, it was expected that similar generalization would occur in this study.

Decision as to which of the categories would be most appropriate for the evaluation of generalization in PREP was based on the following three factors: (1) the sensitivity of the category to the changes observed, (2) the capability of the category to capture the behaviors of importance to the subjects in this particular study, and (3) the accuracy with which the behaviors were recorded within each category

as revealed by the inter observer agrement scores. The selection of the dependent variable was therefore, based in part on a rather circular argument. Category efficacy in measuring generalization was evaluated in terms of the behaviors observed, and therefore, was based on the assumption that these behaviors actually were generalized behaviors. The sensitivity to what changes did exist was seen to be an important factor, and was included in the decision in spite of this dilemma.

In order to look more closely at the separate categories and each one's sensitivity to behavior change with treatment, individual graphs were visually analyzed with respect to baseline stability, data variability, trend change, and level change. Due to the nature of the behaviors being studied, that is, their relatively infrequent occurrence, glaring changes in level and trend were the exception rather than the rule. However, close examination revealed apparent behavior change from baseline scores on several occasions. The results of these analyses, expressed in terms of "yes" there was a change or "no" there did not appear to be a change can be found in Appendix F.

It appeared that any one of the categories of Attending, Object Manipulation, Performance at Last Level of Instruction, the skilled performance score, or the composite score described a change in behavior for the group. Change was evident in 50% or more of all treatments within these categories. Although individual changes were observed in the Touch, Physical Contact, and Performance Below Last Level of Instruction categories, these categories were not the most efficient in demonstrating change for the group of seven subjects. The most

sophisticated levels of performance were among those categories which were most sensitive to change.

Touch

As shown in Figure 2, two of fourteen graphs representing observations described by the Touch category indicated a change in behavior with treatment. Clearly, the frequency of the children's touching behavior did not increase while they received instruction in a skill.

Assuming that some generalization actually did occur in this study and showing that the Touch category did not indicate change for the group, it was concluded that this category was an inadequate tool for describing generalization. In addition, the inter observer agreement scores of 50% and 73% for Observers A and B, indicated that there may have been some problem with the measurement of touching behavior. In light of these points, and the fact that touching behavior did not represent active play behavior, the Touch category seemed inappropriate to describe generalization in this study.

Attend

Eight of fourteen treatment graphs in Figure 3 indicated a positive change in behavior. This suggested that the category of Attending was sensitive enough to capture change in behaviors observed during treatment intervention.

It became apparent after coding for several hours, that a problem existed with the definition of attending as described. Attention was coded any time a child attended to a piece of equipment of instruction, regardless of the skill being performed on it. As a result, skill re-

lated, rather than skill specific behavior may have been measured by this category.

Mean inter observer agreement scores for the Attending category were 59% for Observer A and 76% for Observer B. This suggested that it may have been difficult to discriminate attending behavior while viewing the children on the television monitor. Actually, it was often difficult to see if a child was simply touching a piece of equipment without regard for it or attending to it. This confusion between the Touch and Attend categories may be reflected by the low reliability coefficients in the Touch category also. As a result, the accuracy of attending scores in this study may be limited due to low inter observer agreement scores (Kazdin, 1977).

In addition, scores in the Attending category represented rather passive behavior. A change in the active play of children participating in PREP was not reflected by this category. In light of this fact and apparent measurement problems previously discussed, exclusive use of the Attending category as an indication of generalization in this study seemed inappropriate.

It is conceivable that the measurement problem previously discussed could be lessened if observations were made directly in the classroom rather than on a TV monitor. Assuming that this change might result in accurate measurement, it is possible that this category would be appropriate for use in certain play situations. For example, if a child started out with no play behavior, perhaps a change in attention would be important.

Physical Contact

Inter observer agreement scores calculated on simultaneously coded data indicated that Observers A and B agreed with the criterion observer on 70% and 80% of all behaviors coded in the Physical Contact category. In keeping with Kazdin's (1977) standards for accuracy, it can be assumed that these behaviors adequately represent actual performance.

Although change was demonstrated in five of fourteen treatment graphs presented in Figure 4, the Physical Contact category was not one of the most efficient in demonstrating change for the group. It appeared that the category was not sensitive to changes demonstrated by the subjects during treatment.

Additionally, the behaviors recorded in this category represented passive play behaviors, not necessarily directly related to the skill receiving instruction. That is, the category described only contact with a piece of equipment of instruction, and no intention to initiate the actual skill. Generalization in a program whose goal is to promote active play is not adequately described by this one category.

Object Manipulation

A major change would be recommended for the Object Manipulation category, as originally defined. It became apparent that this category included too many behaviors and in fact, behaviors which were actually quite different from one another. For example, when a child was being taught to ride a scooter and was seen performing any other scooter skill, such as riding a scooter down the incline, it was coded as

Object Manipulation. It was also coded as Object Manipulation if he was observed on his hands and knees, pushing the scooter around the room.

It is recommended that skill-related performance, such as riding a scooter down the incline, be differentiated. It is recommended that skill-related performance be recorded in a separate category in future studies using this instrument. This change would be especially significant to someone investigating response generalization, since a skill related performance is a performance related to, but different from, the instructional skill.

Performance Below Last Level of Instruction

Figure 6 demonstrated that six individual treatment graphs indicated change between baseline and treatment phases in percentage of time spent in the category of Performance Below Last Level of Instruction. In addition to these changes, performance scores for the three treatment A skills for Subject 6 suggest change with treatment. However, baseline instability makes it difficult to evaluate this change.

It appeared that this category was quite sensitive to changes in behavior with treatment. Furthermore, interobserver agreement scores of 86% and 78% for Observers A and B, respectively, imply that this category can be used accurately and objectively.

This category seemed to be an adequate measurement of generalization. In fact, this category might be the best indication of generalization in certain instances. For example, if a child required physical assistance during a considerable number of instructional episodes, there would be no reason to believe that he would perform at a task

step of instruction during free play. It is reasonable to suggest that his skilled behaviors would typically fall into the Performance Below Last Level of Instruction category. The daily record forms for the children in this study however, demonstrated that the children did not require physical assistance for extended lengths of time. On the contrary, skills receiving treatment in the form of maintenance, were never physically prompted.

Exclusive use of this category would fail to consider the range of possible skilled performances. All initiations at and above the task step receiving instruction would be lost. Because the children in this study were often being taught at the visual, verbal, and environmental prompt levels, it was expected that skilled performances by the children would sometimes fall into the Performance At Last Level of Instruction category, and perhaps on occasion in the Performance Above Last Level of Instruction category. For these reasons, exclusive use of this category was considered an inadequate measure of generalization in this study.

Performance At Last Level of Instruction

A conflict arose with the Performance At Last Level of Instruction category. Although each prescribed skill was assessed at a level where the child could actually initiate the skill at some task step, instructional guidelines (Watkinson and Wall, 1980) dictate that instruction begin at the lowest level requiring physical assistance. Therefore, when a child was being taught with one of the three levels of physical prompts, there was no reason to believe that the child would initiate the skill at the task step of instruction. In other words, when a

child needed to be physically prompted to elicit a correct response during a teaching episode, skilled generalization would likely be observed at a task step below the last step receiving instruction. Use of this category by itself to describe generalization of active play behaviors seemed inadequate.

However, use of this category in combination with the Performance Below Last Level of Instruction category might be useful in detecting changes in sophistication of generalized skills. Inter observer agreement scores of 96% and 79% indicated that this category can be used objectively.

Performance Above Last Level of Instruction

It was shown in Figure 8 that skilled performance at a task step higher than the step receiving treatment was seldom observed. Quite clearly, this category was inappropriate for measuring generalization.

Infrequent occurrence in this category revealed an important point. The fact that children seldom performed a skill at a level higher than the task step receiving instruction confirmed that the skill prescriptions had been made correctly. The children were not receiving instruction at a task step they had already learned. Use of this category could serve as a probe for prescription.

Skilled Performance Score

In spite of the individual inadequacy of the separate categories of Performance Below Last Level of Instruction, Performance At Last Level of Instruction, and Performance Above Last Level of Instruction, the combination of all three to represent a skilled performance score

seemed logical for describing generalization of active play behaviors. The skilled performance score required performance of the specific skill receiving treatment. Furthermore, each of the three categories included in the score measured a performance in relation to one thing-the highest task step receiving instruction or maintenance on each day. The first category was included to capture skilled behaviors while a child was being physically prompted during instruction. The Performance At Last Level of Instruction category was included to capture initiated behaviors typically expected once the child reached a level of performance independent of physical assistance. The third category was included to capture the odd performance observed at a level above the last task step receiving instruction. This seemed particularly appropriate for the sessions when a child reached instruction at an environmental prompt level.

A close look at the separate categories of Performance Below, At, and Above Last Level of Instruction, in comparison to the skilled performance score made up of these three scores, revealed an important fact. Behavior changes from one of the first four categories into the more sophisticated categories of Below or At Last Level of Instruction during treatment, would be reflected as an increased initiation from baseline to treatment. In contrast, behavior changes from the Performance Below Last Level of Instruction category into the more sophisticated categories of Performance At or Above Last Level of Instruction, would not be reflected in the percentage scores in the skilled performance graphs. So, the child's behavior could actually become more sophisticated during treatment, yet not be reflected in the graphs.

Therefore, changes observed in Figure 9, describing skilled performance, would be quite conservative in demonstrating change between baseline and treatment performances. And still, the graphs in Figure 9 indicated that generalization seemed to be described in eleven of fourteen treatments. It seemed that this score would be an appropriate tool for measuring change in active free play behavior with treatment.

Just as with the other categories, this may not be the best choice in all situations. For example, use of this score as an instrument to describe generalization would be inappropriate if frequent performances at one of the three levels of skilled performance were observed during baseline, and the attempt was to improve the level of skill used in free play. If this were the case, a change in skilled performance would not be demonstrated by this skilled performance score. Graphs in Figure 6 indicated that this did not restrict the utility of this score in this study.

Finally, the inter observer agreement scores in each of the three categories included in the skilled performance score indicate that they can each be used reliably.

Composite Score

Figures 10a, 10b, and 10c indicated that the composite score made up of all seven category percentage scores, was sensitive to changes demonstrated between baseline and treatment. Ten of fourteen treatment interventions indicated change when the measured behavior was described as any one of the seven defined behaviors.

The categories included in this composite score described a combination of active and passive play behaviors. As mentioned previously,

a change in passive play behaviors such as touching and attention, could be a significant change for a child who sat motionless during initial assessments of free play. However, data indicated that the children in this study spent only 3% of their time in non-play. Because it was the purpose of this study to look at practically significant change in the subjects' behavior, it was felt that inclusion of passive behaviors in the evaluation of generalization was unsuitable.

Decision

Selection of an appropriate category to describe generalization seems to be dependent upon the nature of the questions being asked. In consideration of the questions being asked in this study, the skilled performance score was selected as dependent variable to be used in the evaluation of generalization. Despite problems with the individual categories making up the skilled performance score, the combination of the Performance Below, At, and Above Last Level of Instruction categories was an appropriate trade-off and seemed to be sensitive enough to capture change. Inter observer agreement scores within each of the three categories included in the skilled performance score, indicated that the score could be used to accurately and objectively represent actual performance.

Establishing Criteria for Evaluating Generalization

As it has become more apparent that stimulus generalization is not automatic, studies have begun to demonstrate that certain methods of generalization programming are successful in bringing about generalization (Gable et al, 1978; Jackson and Wallace, 1974). It would seem fair to assume that incorporating these methods into one's program would reasonably improve the chances of seeing skilled performances in situations other than the experimental setting.

Because the PREP Program strategies have been designed to include generalization programming, it is assumed that the chances of seeing stimulus generalization are greater than if attention were not given to such planning. Studies (Noble, 1975; Shatz, 1979; Watkinson, 1977) have shown that PREP strategies improve the quality of play skill performance in young mentally retarded children in the presence of prompts and reinforcement. In addition, Watkinson (1977) demonstrated that children used an increased number of skills in free play after eight months of participation in the PREP Program. However, formal observation of specific skill performance in the absence of pre-response prompting and post-response reinforcement and feedback as an indication of meeting the program goal of stimulus generalization had not been done.

As the investigation of generalization in PREP continued, two new issues arose. First, is graphic analysis an adequate procedure for determining the success or failure of the program? Second, how large a change must be demonstrated before it can be concluded that treatment was effective?

Kratochwill (1978) addressed the first issue by suggesting that behavior changes great enough to be visible to the eye are often greater than the change required to produce statistical significance. A further point in support of graphic analysis or "analysis by inspection" is the fact that graphic data are usually presented in the form of original data and not manipulated by statistical tests. Therefore, readers can make their own judgements about behavior changes rather than relying on a group statistic such as a significant F or non-significant t value.

Evaluation of data using a variety of different graphs led to the identification of the second issue. Traditionally accepted criteria for evaluating graphic data were not appropriate for the data of this study. It seemed that analytical methods based on data properties such as trend and level change were developed for high frequency behaviors, for example, excessive head banging or classes of behavior such as appropriate social behavior. They appeared to be inappropriate for use with low frequency but high value behaviors such as the play skills under investigation.

Since typically acceptable evaluation techniques were inadequate, it became necessary to identify criteria for evaluating change observed in PREP. To a large extent, the question of how large a change is required to produce clear evidence of an effect is left to the discretion of the investigator. Kazdin (1975) suggests that the main criterion in the evaluation of change is in terms of practical significance. In other words, is the behavior change large enough to be of practical importance to the subject? The investigator interprets his data in

terms of his criteria for practical significance. Some will agree with him and others will disagree.

With respect to the question of practically significant behavior change in the data of this study, it seemed that the skills fell into two quite different categories. Some were discrete skills and others usually continued for a short time once initiated. Sulzer-Azaroff and Mayer (1977) differentiate between discrete and continuous skills in the following manner: a discrete skill is a skill with a clearly discriminable beginning and end as opposed to a continuous skill which does not have a clearly discriminable beginning or end. Using this definition, each of the PREP target skills was classified as discrete or continuous (See Table II).

This distinction has implications for the evaluation of generalization within the PREP Program. For example, initiation of the skill of jumping down from an object was usually seen as one episode of jumping down followed by the subject moving on to another activity. On the other hand, when a child chose to ride a scooter he usually rode it for at least a short amount of time.

This differentiation was also especially relevant to the evaluation of results in light of the fact that performance scores were reported as percentages rather than frequency counts. Generally speaking, each interval in which a discrete skill was recorded most often represented one occurrence of the skill. On the other hand, one occurrence of a continuous skill generally lasted longer than one interval.

Percentage scores for discrete and continuous skills, therefore, do not represent similar performances. Initiation of jumping down

Table II: Identification of Discrete and Continuous Skills

Continuous	Discrete
Ascend a Ladder	Jump Over
Jump on the Trampoline	Throwing
Scooter Skills	Swing on a Bar
Hang from the Knees	Scooter on Incline
Swing on the Swing	Kicking
Walk Up an Incline	Forward Roll
Slide Down a Slide	Roll Around a Bar
	Climb Onto a Box
	Jump Down

would have to occur about five times to occupy 1% of free play time available for the day. At the same time, observation of the skill of sitting on a scooter and riding around the room for five intervals would probably describe one interaction with the scooter, and even then a fairly short interaction.

As a result, separate criteria for describing change during and after treatment were specified for discrete and continuous skills. For the discrete skill, a positive change of 1% from baseline to treatment data on 50% of the treatment days was considered to be a significant change between phases. For change to be significant in a continuous skill, 50% of the treatment scores had to be 2% larger than the mean baseline score for that skill.

For example, in Figure 11 for the skill of hanging from the knees, Subject 1 initiated this skill for more time than the mean baseline performance score of .23% during five of seven treatment sessions. Of these five, only Session 9 met the 2% criterion for a continuous skill. Overall criteria for generalization were not met.

An example that meets the criteria for the continuous skill of sliding down a slide can be seen in Figure 12. Percentage scores for Subject 2 were 2% higher than the average baseline score of .45% in four of eight treatment sessions. Since a change was measured in 50% of the treatment sessions, criteria were met and generalization was evident.

For practical purposes these percentage scores can be translated into approximate seconds of performance. Assuming that each child spent 60% of his total observed time in PREP in free play, 1% of

twenty five minutes is calculated to be fifteen seconds. 2% of twenty five minutes is equal to thirty seconds.

These figures may seem unduly liberal to describe a generalized behavior. However, because generalized behaviors were defined as behaviors initiated in the absence of teacher prompts and reinforcers, the child had to chose a skill of instruction instead of a number of other activities and a large variety of available play apparatus. This represents a vast difference from a procedure where a child is prompted and either responds or does not respond, or where overall appropriate behavior is the dependent variable. Quite clearly, the chance of a child engaging in a treated skill is very small.

In addition, the fact that the children spent an average of 35% (with a range of 14% - 62%) of their free play time in teacher interaction, the majority of which was individual instruction in the three prescribed skills, would lead one to suspect that a child might even reach the limit of his tolerance for the skill and rarely chose to initiate it during free play time. Satiation of this kind might be followed by an increased initiation rate during the post treatment phase.

Criteria were set to require performance in only half the sessions because it was felt that this would more closely approximate culturally normative play behavior. It seems unreasonable to expect a child to be interested in the same play activities every day when playing in an environment as varied and stimulating as the PREP room.

In conclusion, a change from zero initiation of a skill to initiating the skill during 1% or 2% of free play time, depending on the particular skill, seemed practically significant. It is proposed that change of this magnitude demonstrated that the PREP Program indeed

had an effect on a child's free play performance.

Generalization of PREP Skills in Individual Subjects

Because the children had each been prescribed six different play skills, comparison between subjects would disregard the possibility that generalization may be skill specific, and that skill prescription rather than skill performance might be responsible for apparent differences. Similarly, evaluation of generalization in terms of treatment A and treatment B effects would overlook the possibility of skill specificity in generalization. Therefore, evaluation of the generalization of PREP skills will be discussed for each individual subject in six individual skills.

While each subject demonstrated a positive change in at least one of six prescribed skills, for Subjects 4 and 5 this change did not meet criteria for even one skill. Subjects 1, 2, 3, and 7 met criteria for one skill, and Subject 6 demonstrated large enough changes in two skills to meet generalization criteria. Daily percentage scores and mean baseline scores for each skill and each subject can be found in Appendix G.

Subject 1

Although Subject 1 performed the skill of hanging from his knees at least 2% more during only one of seven treatment sessions, Figure 11 indicated that he spent greater than 2.23% of his free play time performing the skill in two of three post treatment sessions. This performance indicated a significant change from baseline performance. It also suggested that the child became more interested in initiating

the skill after treatment ended. This data supported comments from the teacher during treatment which suggested that Subject 1 was becoming bored with the skill.

The five remaining graphs in Figure 11 indicated that performance scores did not meet criteria for generalization. Throwing performance in Session 13 fell just short of 1.45% resulting in only one of three sessions indicating change. The fact that Subject 1 was absent from three sessions during the second treatment phase, made it difficult to evaluate the generalization of skills prescribed to him during Treatment B.

Subject 2

Subject 2 increased initiation of sliding down a slide by 2% of his free play time in 50% of eight treatment sessions. This generalization did not continue into the post treatment phase. This return-to-baseline performance could be an indication that Subject 2 needed to be prompted at least initially each day to remind him that sliding down the slide was fun.

Figure 12 clearly indicated that treatment in five additional skills failed to result in generalized performance in free play.

Subject 3

Subject 3 met the 2% criterion for the skill of scooter riding in every treatment session and therefore generalization of this skill was unquestionable. Similar performance in other skills was not observed. In fact, only one other skill for Subject 3 even indicated slight generalization.

When skills were classified as discrete or continuous, the skill of sliding down a slide appeared to fall somewhere in the middle of the two classifications. The decision was made to classify sliding as a continuous skill, because many of the children were observed sliding down the slide, particularly on their tummy, and stopping part way down for a short time. Had it been identified as a discrete skill, Subject 3 would have met the 1% increase over baseline performance criterion in five of the seven sessions. Therefore, although generalization on the basis of the identified criteria was not evident, a fairly significant change was demonstrated.

Subject 4

Although graphs in Figure 14 indicated that Subject 4 initiated the skills prescribed to him during treatment, the percentage of free play time spent performing them did not meet criteria for even one skill.

Performance in jumping on the trampoline was maintained above zero baseline performance in eight of eleven treatment and post treatment sessions. The percentage score of 7.6% in Session 11 represented the only significant change from the mean baseline performance. While the data for this subject suggested a change in initiation of jumping on the trampoline during and after treatment, they did not indicate generalization as defined in this study.

Subject 5

Subject 5 did not meet criteria for generalization in any skill. The first graph in Figure 15 appeared to indicate significant

change in the skill of climbing up a ladder. Looking more carefully at each percentage score, this child spent 2% more time than the average baseline score of zero percent on only one occasion. While increased initiation was apparent, the change was not large enough to be considered a practically significant change in Subject 5's free play behavior.

Similarly, slight and occasional change was indicated for the skills of jumping over, jumping on a trampoline, and scooter riding. However, the change from baseline data was not large enough to justify generalization for these skills.

Subject 6

Free play behavior observation for Subject 6 indicated that generalization occurred for two of six prescribed skills. Although Subject 6 did not meet criteria for sliding during the treatment sessions, she did spend greater than 4.2% of her free play time initiating the skill in three of five post treatment sessions. Once again, this behavior pattern suggested that the child became more interested in initiating the skill in free play after treatment stopped.

Data for the skill of ascending a ladder in Figure 16 demonstrated that Subject 6 spent more than 2.5% of her free play time performing the skill while receiving instruction, 80% of the time. Quite clearly, this met the criteria for generalization.

It is interesting to note that these two skills of ascending a ladder and sliding strongly influence the initiation of each other. As the possibility of Subject 6 jumping down from the five foot high climbing apparatus at the top of the ladder was nearly zero, she had

little choice but to slide down the slide after climbing up the ladder. While it is not possible to say which skill was more important in motivating performance of the other, it suggested that teaching pairs of related skills may actually increase generalization performance of the skills in free play.

In addition, free play performance of Subject 6 in swinging on a bar gradually improved during treatment. Because the decision to stop treatment A and begin Treatment B was made on the basis of group performance, treatment in this skill may have been stopped before it should have for Subject 6. Percentage scores in Sessions 10 and 11 indicated significant change for this discrete skill, suggesting that generalization was occurring, but only after several treatment sessions.

Subject 7

Free play performance graphs in Figure 17 showed that Subject 7 spent a considerably larger percentage of his time walking up an incline during treatment than during baseline. In fact, four of seven treatment session scores were greater than 2%. Observation of the skill being initiated in free play continued into the post treatment phase however, the percentage scores dropped below the 2% criterion in four of six post treatment sessions.

Data in the ladder climbing and throwing graphs suggested a considerable change in performance once treatment began. However, scores for the skill of ascending a ladder did not meet criteria, and throwing performance exceeded the mean baseline score by 1% during one treatment session and one post treatment session. Therefore, it can only be concluded that generalization of the skills of throwing and climbing

a ladder was observed in free play, but not often enough to suggest that PREP intervention significantly changed the behavior.

Figure 17 indicated that skills treated in the second treatment phase showed very little sign of generalization. In fact, Subject 7 was never observed swinging on a bar.

It is interesting to look at the graph for riding on a scooter. Initiation of this skill was not observed until the fifth day of treatment. During the last two treatment sessions, performance scores began to improve. Although the school year finished and the program ended, this trend would seem to suggest that further treatment sessions might have resulted in generalization. Data such as that in the scooter riding graph in Figure 17 suggest that treatment length should be prescribed on an individual basis rather than for the group as a whole.

Summary of Generalization in Individual Subjects

The data of this study indicated that five of seven subjects met the criteria for generalization in at least one of six skills. Consequently, program success in terms of meeting the goal of stimulus generalization was somewhat limited.

A variety of factors may have influenced the failure of the individually prescribed skills to become intrinsically motivating to the children. In retrospect, two particular explanations seem especially important to the findings of this study. These include the influence of competing preferential skills and the adequacy of reinforcement procedures to increase low frequency behaviors.

First, consider the issue of preferential skills. While this in-

vestigation attempted to increase the frequency of learned behaviors which were rarely initiated, the results indicated that the treatment did not have a considerable effect on these behaviors. The competition of highly valued skills may have been overpowering. In other words, the number of highly motivating skills in the children's free play repertoires at the start of the study seemed to be adequate to keep them busy during their free play time in the PREP playroom.

Secondly, the possibility exists that the reinforcement procedures identified in the PREP Program are inadequate to increase the frequency of these low frequency skills. This is an important consideration in light of the fact that the skills chosen for treatment were initially assessed at a level of independence. Although the children had learned the skills previously, naturally reinforcing contingencies were not strong enough to prompt their use in free play. It is possible that a revision in reinforcement procedures to provide more reinforcement as the child progressed through the task sequence may enhance generalization.

Skill Specificity of Generalization

While five subjects demonstrated generalization during and/or after treatment in at least one skill, many skills did not generalize. Analyses of the initiation rates of a variety of skills prescribed to the subjects indicated that generalization of certain skills may be more pronounced than others. In order to look at this more closely, individual performance records for skills prescribed to three or more subjects were grouped together. The assumption was that repeated evidence of generalization across subjects strengthens the predictability of specific skill generalization.

The results of individual skill performance might have implications for skill prescription guidelines. For example, if a skill fails to generalize according to criteria in all five subjects prescribed the skill, is it worthwhile spending valuable teaching time to improve performance during instruction? If some skills are more apt to be initiated in the absence of prompts and reinforcement than others, perhaps effort should be concentrated on those, particularly with children who have few skills in their free play repertoire.

Ascend a Ladder

As shown in Figure 18, change in performance with treatment occurred in all children taught to ascend a ladder. Data for Subject 6 met criteria for this continuous skill. Scores for Subject 5 failed to meet criteria but increased at least 1% in half of all treatment and post treatment session. Baseline instability made it difficult to evaluate change as a result of treatment for Subjects 4 and 7.

On the basis of performance records of four children receiving instruction or maintenance on the skill of ascending a ladder, generalization as described was evident 25% of the time.

Jump on the Trampoline

Since trampoline jumping was classified as a continuous skill, data had to meet the criterion of 2% change in time spent jumping on the trampoline during 50% of all treatment sessions to show evidence of generalization. While baseline instability for Subject 5 made it untenable to draw conclusions about change between phases, data in Figure 19 for Subject 3 show clearly that generalization failed to

occur.

Informal observation of free play behavior revealed that the children often stood on the trampoline without jumping. This type of behavior would be coded in the Object Manipulation category rather than the Performance Below, At, or Above Last Level of Instruction categories. This behavior would not be reflected in the skilled performance score. Because of the nature of this skill, it is therefore suggested that the change shown for Subject 4 indicated a practically significant change. He was observed actually jumping during five of seven sessions, but he was not spending great amounts of time at it. Further instruction may have resulted in additional data similar to Session 11. However, generalization was not indicated by the available data on the basis of identified criteria.

In summary, the skill of jumping on a trampoline generalized as a result of treatment in one of three instances.

Jump Over

Graphs presented in Figure 20 indicate obvious failure for the skill of jumping over to generalize from instructional situations into free play. Lack of generalization during and after treatment was demonstrated for all five subjects receiving maintenance or instruction in this skill. The data quite clearly showed that children seldom chose to jump over an object for the fun of it.

Scooter Riding on the Floor

Four subjects were prescribed the skill of riding a scooter on the floor while sitting on it. Performance graphs are presented in Figure

21. Subject 3's performance very clearly demonstrated generalization. Subject 5 spent increased amounts of free play time riding a scooter during the initial treatment sessions however, performance returned to baseline zero rate during Sessions 16 and 17. Although a definite change in performance was evident with the onset of treatment, the fact that changes did not meet criteria, and improvement gradually faded made it difficult to justify generalized effects.

Subject 6 never initiated this scooter skill in free play and Subject 4 used it during only one treatment session. Although generalization for scooter riding was demonstrated without question for Subject 3, the three additional subjects receiving treatment performed quite differently. It appeared that generalizability was not predictable for this skill.

Sliding Down a Slide

Looking at Figure 22, it seemed fair to suggest that generalization of sliding is somewhat predictable. All three graphs met generalization criteria during the treatment phase. In addition to meeting criteria, data for Subject 7 demonstrated a change in trend and level in the treatment phase. Change could be suitably described as a level change for Subject 3. This change was somewhat decaying (Kratochwill, 1978) but remained above baseline performance rates during the three post treatment sessions.

The graph for Subject 6 in Figure 22 indicated a definite trend change between baseline and post treatment data. Although the child initiated the skill during every treatment session, daily percentage scores rose above the mean baseline score only once. Post treatment

data paired with the initiation during each treatment session strongly suggested generalized treatment effects. However, Subject 6 also received instruction in ladder climbing during the second treatment phase, therefore it is inappropriate to propose that treatment of the skill of sliding was the only factor affecting her sliding performance. At the same time, one cannot discount the fact that she was observed using the skill.

Swing on a Bar

Performance records for four subjects for swinging on a bar were shown in Figure 23. In two of four instances the subject never initiated the skill in free play during treatment. A third subject, Subject 4, was observed using the skill a small amount (less than 1%) in two of eight treatment sessions. Graphs for Subjects 3, 4, and 7 demonstrated failure of the skill to generalize.

Data for Subject 6 did not meet criteria for generalization. However, data for Subject 6 indicate a trend change during treatment. The change could be described as a delayed change in trend (Kratochwill, 1978). The performance graph for this subject again indicated that additional instruction may have improved generalization of the skill in free play.

In looking at the four graphs in Figure 23 it seemed that more often than not, children taught the skill of swinging on a bar, did not spend much of their free play time practicing the skill.

Throwing

The skill of throwing a ball was prescribed to four subjects. As shown in their performance graphs in Figure 24, each of these children were observed using the skill in free play to a certain extent. Data for Subject 7 was most convincing in demonstrating change as a result of treatment.

An important point here is that children typically would throw a ball to another child in a catching game. Failure to find a partner skilled enough to play in this manner in the PREP Program might act to reduce the use of this skill in free play. Interestingly, Subject 1 was observed calling a teacher over to play catch with him on more than one occasion, indicating that he recognized the need for a skilled partner. This behavior would be coded as Teacher Interaction and would not be reflected in the generalized data. Although the success of actually teaching social interaction skills in PREP is limited (Wasson, 1980), it is recommended that throwing may be a skill which would best generalize after instruction between pairs of children.

Summary of Skill Specificity in Generalization

Analysis of specific skill generalization across subjects in this study indicated that generalization of some skills may be more pronounced than others. To a certain extent, generalization may be skill specific.

Generalization was not demonstrated by any of the children in the skills of jumping over, swinging on a bar, and throwing. While it can not be concluded that these skills will not generalize with other children, it was a significant finding that none of the children in

this study were intrinsically motivated to perform these skills. This information could act as a guideline in the prescription of skills in PREP and similar instructional programs. Valuable instructional time may be better spent teaching skills which are more likely to be reflected in the children's free play performance.

For each of the skills of ascending a ladder, jumping on the trampoline, and riding a scooter on the floor, generalization was demonstrated by one child. Since these skills were taught to at least three children, generalization was shown to be the exception rather than the rule. Predictability of generalization in these three skills is limited. However, the results of this study indicated that instruction in these skills may significantly change the free play behavior of some children.

Analysis of skill generalization across subjects indicates that generalization of one of seven skills taught to three or more children may be predictable. Performance graphs for all three children prescribed the skill of sliding down a slide met criteria for generalization. This is especially relevant in light of the fact that only six of forty two treatments produced generalized effects. It is suggested that instruction in the skill of sliding will more often than not result in a practically significant change in the initiation of this skill in free play.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Summary and Conclusions

The ultimate goal of this research was to evaluate the stimulus generalization of motor skills in the PREP Program. Sixty minutes of free play behavior was videotaped during seventeen sessions, providing the raw data for the study. Eight participants in the PREP Program were initially selected as subjects, however, one subject was deleted due to frequent absences.

Six skills were prescribed for each subject from the list of PREP target skills. Each of these skills had been assessed at a level of independence, and therefore, could be initiated during free play time without any assistance from the teacher. Two of the six skills were prescribed for instruction and four were prescribed for maintenance. Skills were treated according to instructional and maintenance guidelines specified for this study, revised from PREP instructional strategies. The prescribed skills were divided into two groups. Three skills were treated during the eight sessions of Treatment A, and three new skills received treatment in six sessions of Treatment B.

Because identification of qualitative and quantitative criteria for the assessment of generalization has received inadequate attention, a category observation instrument was designed to capture the range of behaviors related to generalized skill performance. The investigator and two trained observers utilized the instrument to transcribe the information provided on the videotape recordings into descriptive free

play data for the seven subjects. The descriptive data collected, represented an accurate and objective measure of the behaviors of importance, as revealed by the 85% and 82% inter observer agreement scores.

Selection of a dependent variable to measure generalization was based on the sensitivity of separate categories to the changes observed, the capability of each category to capture the active play behavior of the seven children included in this study, and the reliability of each category. An extensive analysis eventually led to the selection of a skilled performance score as dependent variable. This skilled performance score included the total percentage of free play time recorded in the Performance Below Last Level of Instruction, Performance At Last Level of Instruction, and Performance Above Last Level of Instruction categories. Although this score was chosen as the most appropriate measure of generalization for this study, the possibility of using different categories in other situations was discussed.

Criteria for describing behavior change during and after treatment were identified because typically acceptable techniques for evaluating graphic data were inadequate for the low frequency yet vitally important behaviors under study. Separate criteria were specified for discrete and continuous skills.

Close examination of performance graphs for individual subjects, with respect to these criteria, revealed that some generalized behavior was observed for all subjects. The extent of change was variable across subjects. Four subjects met criteria for generalization in one skill. One subject met criteria in two skills, and two subjects did not meet criteria in any of the skills prescribed to them.

Analysis of the generalization of specific skills, based on performances of the seven skills prescribed to at least three children, indicated that generalization of some skills may be more pronounced than others. It was concluded that generalization may be somewhat skill specific.

The evaluation of generalization in this study was based on the following assumptions: (1) that the skilled performance score including percentage scores in the Performance Below Last Level of Instruction, Performance At Last Level of Instruction, and Performance Above Last Level of Instruction categories, adequately described generalized behavior, and (2) specified criteria were reasonable for assessing generalization of each of the skills in relation to culturally normative play patterns. The first assumption seems to be a safe one in light of past research. The second assumption is justified on the logic that criteria were specified with special consideration of typically observed free play behavior in young moderately mentally retarded children.

Recommendations for Further Research and Programming

A number of recommendations are suggested for future research of a similar nature to this study. Additionally, several seemingly important points in regards to PREP programming became apparent over the duration of this investigation.

(1) Formal observation of free play patterns of young healthy children of average intellect could result in the establishment of norms describing typical amounts of time spent in PREP target skills during free play. Information revealed by a study of this nature would

affirm whether the criteria used to evaluate generalization in this study was reasonable.

(2) Special consideration should be given to the number of skills in a child's free play repertoire when evaluating the generalization of new skills or initially low frequency behaviors. The influence of competing preferential skills should be included in the evaluation.

(3) The dilemna of when to terminate treatment is an important issue in the PREP Program. As indicated by the data and teacher comments, there seems to be a trade-off between maintaining teacher and student enthusiasm and the length of instruction in a skill. Data in this study indicate that termination of treatment should be evaluated on an individual basis in order to maximize use of instructional time.

(4) On the basis of the skill performance of Subject 6, it is recommended that pairs of related skills such as sliding and climbing a ladder, throwing and catching, and climbing onto a box and jumping down, can be taught in an effort to maximize generalization of skills.

(5) Some of the children in this study increased the amount of time spent initiating prescribed skills after treatment ended. This could be interpreted as evidence of satiation. It is therefore recommended that instructional strategies could include teaching skills alternately for short periods of time. For example, a teacher could teach one skill for four sessions, teach a second skill for four sessions, and then return to teach the first skill again. The data of this study indicate that this type of teaching schedule may enhance generalization.

(6) It is recommended that teacher assignments in PREP be planned so that teachers teach several different children one skill each, rather

then several skills to one child. Teacher comments indicated that they enjoyed teaching a variety of children.

(7) Reinforcement procedures used in PREP may need alteration, particularly when children receive instruction in previously learned but seldom initiated play skills. Perhaps heavy reinforcement should be introduced at every task step once the child reaches a verbal cue level, before moving on to completely manipulate him at the next higher task step.

(8) Finally, the importance of viewing videotaped recordings of activity in the PREP Program can not be too strongly impressed. Information provided by these records allow teachers to see a wide variety of situations ordinarily overlooked. This recommendation has particular implications for teacher training in the PREP Program.

REFERENCES

Arrington, R. E. Time sampling studies of child behavior. Psychological Monographs, 1939, 51, 2, 1 - 194.

Baer, D. M., Wolf, M. M., and Risley, T. R. Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1968, 1, 91 - 97.

Bateman, S. Application of Premack's generalization on reinforcement to modify occupational behavior in two severely retarded individuals. American Journal of Mental Deficiency, 1975, 79, 5, 604 - 610.

Bijou, S. W., Peterson, R. F., Harris, F. R., Allen, K. E., and Johnston, M. S. Methodology for experimental studies of young children in natural settings. The Psychological Review, 1969, 19, 177 - 210.

Birnbrauer, J. S., Peterson, C. R., and Solnick, J. V. Design and interpretation of studies of single subjects. American Journal of Mental Deficiency, 1974, 79, 2, 191 - 203.

Buell, J., Stoddard, P., Harris, F.R., and Baer, D. M. Collateral social development accompanying reinforcement of outdoor play in a preschool child. Journal of Applied Behavior Analysis, 1968, 1, 167 - 173.

Campbell, D. T. and Stanley, J. C. Experimental and Quasi-Experimental Designs for Research, Chicago: Rand McNally College Publishing Company, 1963.

Connolly, K. Ecological techniques and the direct observation of behavior. In: P. Mittler (Ed.), Assessment for Learning in the Mentally Handicapped, Edinburgh: Churchill Livingstone, 1973, 219 - 245.

Cooke, T. P. and Appoloni, T. Developing positive social-emotional behaviors: a study of training and generalization effects. Journal of Applied Behavior Analysis, 1976, 9, 1, 65 - 78.

Cuvo, A. J. Multiple baseline designs in instructional research: pitfalls of measurement and procedural advantages. American Journal of Mental Deficiency, 1979, 84, 3, 219 - 228.

Cuvo, A. J., Leaf, R. B., and Borakove, L. S. Teaching janitorial skills to the mentally retarded: acquisition, generalization, and maintenance. Journal of Applied Behavior Analysis, 1978, 11, 345 - 355.

Ellis, M. J. Play: a paradox for teacher and scientist. Quest, 1976, 26, 128 - 138.

Flanders, N. A. Interaction Analysis in the Classroom, University of Michigan, 1966.

Flavell, J. E. Reduction of stereotypies by reinforcement of toy play. Mental Retardation , 1973, 11, 21 - 23.

Gable, R. A., Hendrickson, J., and Strain, P. S. Assessment, modification, and generalization of social interaction among severely retarded, multihandicapped children. Education and Training of the Mentally Retarded, 1978, 3, 279 - 286.

Guralnick, M. J. and Kravik, M. A. Reinforcement procedures and social behaviors in a group context with severely retarded children. Psychological Reports, 1973, 32, 295 - 301.

Hall, R. V. and Broden, M. Behavior changes in brain-injured children through social reinforcement. Journal of Experimental Child Psychology, 1967, 5, 463 - 479.

Hanley, E. M. Review of research involving applied behavior analysis in the classroom. Review of Educational Research, 1970, 40, 5, 597 - 624.

Hardiman, S. A., Goetz, E. M., Reuter, K. E., and LeBlanc, J. M. Primes, contingent attention, and training: effects on a child's motor behavior. Journal of Applied Behavior Analysis, 1975, 8, 399 - 409.

Hawkins, R. P. and Dotson, V. A. Reliability scores that delude: an Alice in Wonderland trip through the misleading characteristics of inter-observer agreement scores in internal recording. In: E. Ramp and G. Semp (Eds.), Behavior Analysis: Areas of Research and Applications, Englewood Cliffs, New Jersey: Prentice Hall, 1975, 359 - 376.

Horne, B. M. and Philleo, C. C. A comparative study of spontaneous play activities of normal and mentally defective children. Journal of Genetic Psychology, 1942, 61, 34 - 46.

Hutt, S. J. and Hutt, C. Direct Observation and Measurement of Behavior, Springfield, Illinois: Charles C. Thomas Publisher, 1974.

Jackson, D. A. and Wallace, R. F. The modification and generalization of voice loudness in a fifteen-year-old retarded girl. Journal of Applied Behavior Analysis, 1974, 7, 461 - 467.

Jensen, P. K. Increasing cooperative social interactions between kindergarten children in a free play setting. Unpublished Doctoral Dissertation, Department of Physical Education, University of Alberta, 1979.

Johnson, S. M. and Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In: Hamerlynck, L. A., L. C. Handy, and E. J. Mash, Behavior Change: Methodology, Concepts, and Practice, the Fourth Banff International Conference on Behavior Modification, Research Press, Champaign, Illinois, 1973.

Johnston, M. K., Kelley, C. S., Harris, F. R., and Wolfe, M. M. An application of reinforcement principles to development of motor skills of a young child. Child Development, 1966, 37, 379 - 387.

Kazdin, A. E. Artifact, bias, and complexity of assessment: the ABC's of reliability. Journal of Applied Behavior Analysis, 1977, 10, 1, 141 - 150.

Kazdin, A. E. and Bootzin, R. R. The token economy: An evaluative review. Journal of Applied Behavior Analysis, 1972, 5, 343 - 372.

Keeran, C. V., Grove, F. A., and Zachofsky, T. Assessing the playground skills of the severely retarded. Mental Retardation, 1969, 7, 29 - 32.

Kratochwill, T. R. Single Subject Research: Strategies for Evaluating Change, New York: Academic Press, 1978.

Langone, J. and Westling, D. L. Generalization of prevocational and vocational skills: some practical tactics. Education and Training of the Mentally Retarded, 1979, 14, 3, 216 - 221.

Levy, J. Play Behavior, New York: John Wiley and Sons, 1978.

Linford, A. G., Jenrenaud, C. Y., Karlsson, K., Witt, P., and Linford, M. D. A computerized analysis of characteristics of Down's Syndrome and normal children's free play patterns. Journal of Leisure Research, 1971, 3, 44 - 52.

Martin, J. Developing category observation instruments for the analysis of classroom behavior. Journal of Classroom Interaction, 1976, 12, 5 - 16.

Noble, A. An instrument to measure the sensori- motor play of preschool trainable mentally retarded children. Unpublished Masters Thesis, Department of Physical Education, University of Alberta, 1975.

Peterson, N. L. and Haralick, J. G. Integration of handicapped and non-handicapped preschoolers: An analysis of play behavior and social interaction. Education and Training of the Mentally Retarded, 1977, 12, 3, 235 - 245.

Pierce, C. H. and Graland, D. Effects of four training procedures on increasing motor skills of six physically handicapped retarded persons. Journal of Mental Deficiency Research , 1977, 21, 1, 25 - 35.

Rincouver, A. and Koegel, R. L. Setting generality and stimulus control in autistic children. Journal of Applied Behavior Analysis, 1975, 8, 235 - 246.

Shatz, D. M. An implementation and evaluation of the PREP Progam in a rural community. Unpublished Masters Thesis, Department of Physical Education, University of Alberta, 1979.

Stokes, T. F. and Baer, D. M. An implicit technology of generalization. Journal of Applied Behavior Analysis, 1977, 10, 2, 349 - 367.

Stokes, T. F., Baer, D. M., and Jackson, R. L. Programming the generalization of a greeting response in four retarded children. Journal of Applied Behavior Analysis, 1974, 7, 599 - 610.

Strain, P. Increasing social play of severely retarded preschoolers with socio-dramatic activities. Mental Retardation, 1975, 13, 6, 7 - 9.

Strain, P. S. and Shores, R. E. Additional comments on multiple-baseline designs in instructional research. American Journal of Mental Deficiency, 1979, 84, 3, 229 - 234.

Sulzer-Azaroff, B. and Mayer, G. R. Applying Behavior-Analysis Procedures with Children and Youth, New York: Holt, Rinehart, and Winston, 1977.

Terry, B. Development and use of a behavior observation instrument. Unpublished paper, Department of Physical Education, University of Alberta, 1980.

Tilton, J. and Ottinger, D. P. Comparison of toy play behavior of autistic, retarded, and normal children. Psychological Reports, 1964, 15, 967 - 975.

Walker, H. M. and Buckley, N. K. Programming generalization and maintenance of treatment effects across time and settings. Journal of Applied Behavior Analysis, 1972, 5, 209 - 224.

Walker, H. M., Hops, H., and Johnson, S. M. Generalization and maintenance of classroom training effects. Behavior Therapy, 1975, 6, 188 - 200.

Wall, A. E. Equipment preference and free play practice patterns of preschool moderately mentally retarded children. Unpublished paper, PREP Program, Department of Physical Education, University of Alberta, 1974.

Watkinson, E. J. The effect of the PREP preschool play program on the play skills and free play patterns of moderately mentally retarded children. Unpublished Doctoral Dissertation, Michigan State University, 1977.

Watkinson, E. J., Jordan, A. E., and Wall, A. E. A model for the observation and analysis of teacher behaviors in individualized instruction with mentally retarded children. Unpublished paper, PREP Program, Department of Physical Education, University of Alberta, 1980.

Watkinson, E. J. and Wall, A. E. The PREP PLAY PROGRAM: Play Skill Instruction for Young Mentally Retarded Children, Department of Physical Education, University of Alberta, 1980.

Wehman, P., Abramson, M., and Norman, C. Transfer of training in behavior modification programs: an evaluative review. Journal of Special Education, 1977, 11, 2, 217 - 231.

Wehman, P. and Marchant, J. A. Improving free play skills of severely retarded children. American Journal of Occupational Therapy, 1978, 32, 2, 100 - 104.

Weininger, O. Unstructured play as a vehicle for learning. International Journal of Early Childhood, 1972, 4, 2, 63 - 69.

Whitman, T. L., Mercurio, J. R., and Caponigri, V. Development of social responses in two severely retarded children. Journal of Applied Behavior Analysis, 1970, 3, 2, 133 - 138.

APPENDIX A

SUBJECTS AND PRESCRIBED SKILLS

M - Maintenance
I - Instruction

Subject 1	I - Hang from Knees M - Jump Over M - Swing on Swing
	I - Roll Around a Bar M - Forward Roll M - Throwing
Subject 2	I - Roll Around a Bar M - Slide Down a Slide M - Jump Over
	I - Hang from Knee M - Scooter Riding Down Incline M - Forward Roll
Subject 3	I - Jump on Trampoline M - Jump Over M - Slide Down Slide
	I - Swing on a Bar M - Throwing M - Ride Scooter on Floor Sitting
Subject 4	I - Ascend a Ladder M - Jump Over M - Swing on a Bar
	I - Jump Down M - Kicking M - Ride Scooter on Floor Sitting
Subject 5	I - Ascend a Ladder M - Jump Over M - Jump on a Trampoline
	I - Throwing M - Jump Down M - Ride Scooter on Floor Sitting

M - Maintenance
I - Instruction

Subject 6	I - Swing on Bar M - Slide Down a Slide M - Jump Over	
	I - Ascend a Ladder M - Kicking M - Ride Scooter on Floor Sitting	
Subject 7	I - Ascend a Ladder M - Walk Up an Incline M - Jump on Trampoline	
	I - Throwing M - Jump Down M - Ride Scooter on Floor Sitting	

APPENDIX B

CODING RULES

General Coding Rules

1. Code the most sophisticated behavior category if a behavior fits into more than one category.
2. When more than one behavior related to a prescribed skill is observed in a three second interval, code the most sophisticated behavior.
3. When behaviors related to more than one prescribed skill are observed, code them both.
4. Code a behavior as initiated unless a definite teacher interaction is observed.
5. When a child moves out of camera view but part of the child's body or the teacher can still be seen, code same as last category or the most logical category.
6. Code behavior as Physical Contact when there is regard for the equipment, that is;

   (a) equipment is meant to be manipulated in the hands and it is held in the hands.

   (b) equipment is meant to support the body weight and it is supporting the body.
7. Code behavior as Touches where there is no regard for the equipment.
8. Code being in a "special place" as playful.
9. If a child receives assistance from another child on a prescribed skill, code it as Below Last Level of Instruction.
10. Consider the large slide and orange padded slide as slides, code all other inclines as a Performance Below Last Level of Instruction.
11. When a child watches another child's instructional episode on one of his prescribed skills, code it as attending.

12. When the cue to begin a new row in coding a segment is heard and the scorer is one space behind, skip it and begin the new row.
13. When the cue to begin a new row is sounded and the scorer is one space ahead, scratch the first one out and begin the new row again.

Skill Specific Coding Rules

1. Child is being taught to hang from his knees from a ladder,

   (a) seen sitting on top of ladder - Object Manipulation
   (b) seen swinging on a rung - Object Manipulation
   (c) seen swinging and kipping legs up to rung - Performance Below Last Level of Instruction

2. Child is being taught to jump over a hoop,

   (a) seen carrying a hoop - Playful

3. Child is being taught to ascend a ladder,

   (a) seen descending a ladder - Object Manipulation

4. Child is being taught any scooter skill,

   (a) any other scooter skill - Object Manipulation

5. Child is being taught to ride a scooter on his seat,

   (a) observed pushing the scooter on his hands and knees - Object Manipulation

6. Child is being taught to jump on a trampoline,

   (a) seen sitting on trampoline attending to another child jumping on the trampoline - Object Manipulation
   (b) seen sitting on the trampoline alone - Object Manipulation
   (c) seen trying Task 1 of the seat drop target skill - Object Manipulation

7. Child is being taught to slide down a slide,

   (a) seen sitting at the top of slide - Object Manipulation
   (b) slides down and sits at bottom of slide - Object Manipulation for 3 intervals and then Physical Contact
   (c) walking up to a slide and lying down at the bottom - Physical Contact
   (d) seen ascending a ladder on the big platform - Playful
   (e) seen playing on the big platform - Playful

8. Child is <u>not</u> being taught to slide down a slide,

   (a) seen waiting at the top of the slide - Playful

9. Child is being taught to slide down a slide head first,

   (a) seen pulling himself along a level incline - Object Manipulation

10. Child is being taught to kick a ball,

    (a) seen carrying a ball - Playful
    (b) seen throwing a ball - Playful

11. Child is being taught to throw a ball,

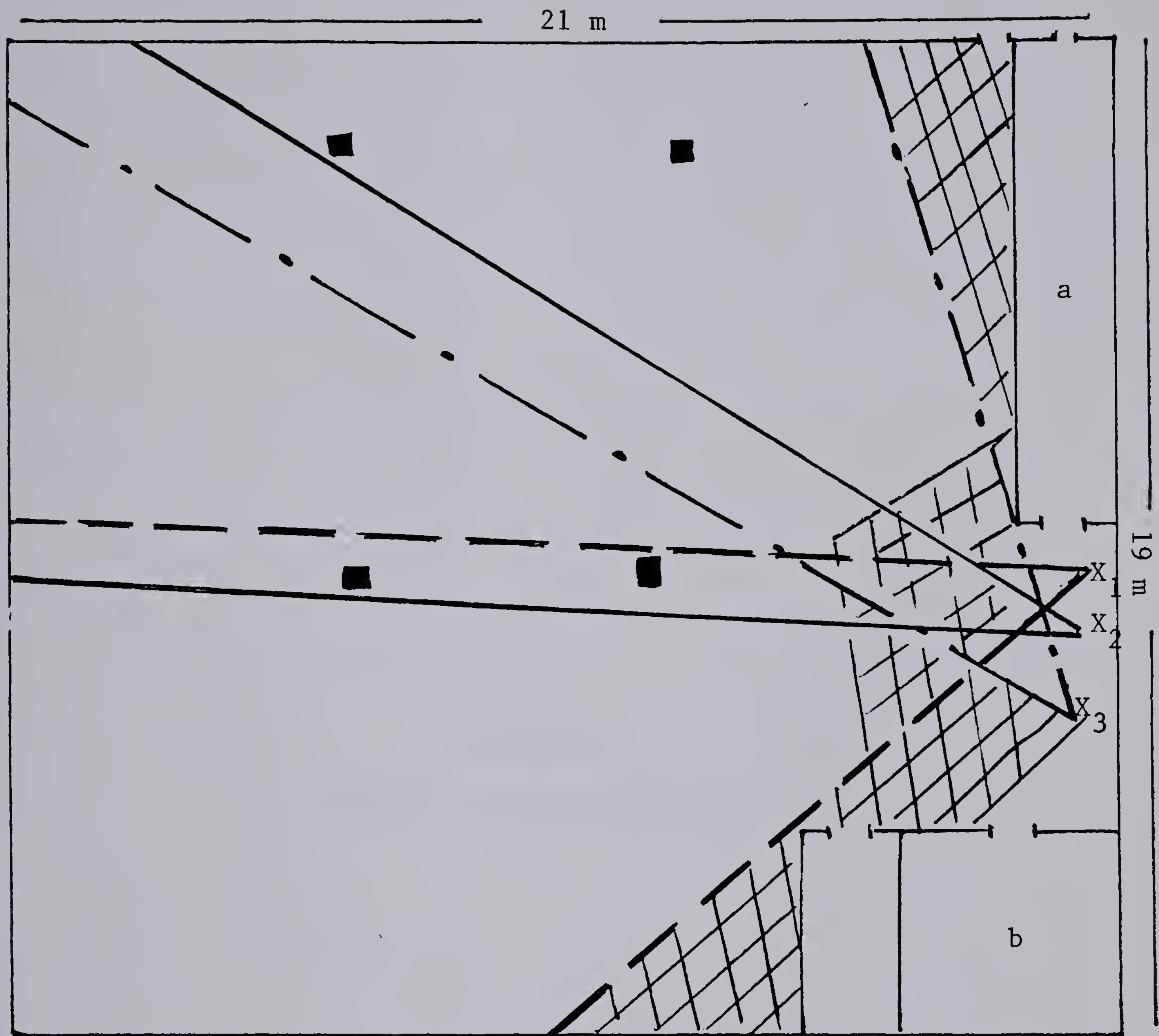
    (a) seen kicking a ball - Playful
    (b) seen carrying a ball - Object Manipulation
    (c) seen sitting and holding a ball - Physical Contact
    (d) seen playing baseball - Attends, when the other child throws

12. Being taught to swing on a bar,

    (a) seen standing on a bench and holding a bar with two hands - Object Manipulation
    (b) seen standing on floor and holding a bar with two hands - Physical Contact
    (c) seen sitting on floor and holding bar with two hands - Physical Contact
    (d) seen sitting on floor and holding bar with one hand - Touches
    (e) seen sitting on top of the low bar - Object Manipulation
    (f) seen hanging from hands and knees on a bar - Object Manipulation
    (g) seen watching someone else roll around a bar - Attends

APPENDIX C

OPTIMAL CAMERA PLACEMENT



support pillar

area not covered by any camera

X - camera

a - observation room with one way mirror

b - office

APPENDIX D

CATEGORY SYMBOLS FOR CODING

Playful	P
Non-Playful	NP
Teacher Interaction	Ti
Unobservable	U
Touches	T
Attends	A
Physical Contact	PC
Inappropriate Object Manipulation	IO
Object Manipulation	O
Performance Below Last Level of Instruction	B
Performance At Last Level of Instruction	At
Performance Above Last Level of Instruction	Ab

APPENDIX E

TRAINING AGREEMENT QUIZ

TRAINING AGREEMENT QUIZ

Please indicate the correct behavior category for the following play behaviors. Assume that each of the skills is being instructed or maintained at target skill level. Assume that the equipment mentioned is being used in instruction or maintenance unless stated otherwise.

_______________ 1. Pushing scooter around with hands while on knees.

_______________ 2. Riding scooter down incline, legs and arms lifted.

_______________ 3. Riding scooter down incline, feet dragging.

_______________ 4. Watching a child ride down incline, clapping hands.

_______________ 5. Sits in a cube and chats with someone in next cube.

_______________ 6. Sitting stationary on a swing.

_______________ 7. Using two scooters on floor, one to support tummy and one to support legs.

_______________ 8. Child moves behind a post for more than three seconds.

_______________ 9. Sitting on a hockey stick.

_______________ 10. Watching an unoccupied swing.

_______________ 11. Tummy on scooter on floor, pulls and glides.

_______________ 12. Throws a hockey stick.

_______________ 13. Brushes a cube as he walks by.

_______________ 14. Using a hockey stick as a bat.

_______________ 15. Using a hockey stick to hit someone.

_______________ 16. Sits on a scooter but doesn't move.

_______________ 17. Sits on scooter on floor, pushes with both feet.

_______________ 18. Walks around swinging a hockey stick.

_______________ 19. Pushes someone on a swing.

_______________ 20. Sits at bottom of slide.

_______________ 21. Swings on swing and pumps legs.

_______________ 22. Lies down on slide and remains stationary.

_______________ 23. Carries scooter in hands, as walks up incline.

______________ 24. Pushes scooter up incline with hands.

______________ 25. Sits on scooter and follows in a train.

______________ 26. Lies down on empty trampoline, no trampoline skills.

______________ 27. Stands on scooter and rides down incline.

______________ 28. Child not in view.

______________ 29. Sits on mat and stares up at ceiling.

______________ 30. Steps on hockey stick while walking around room.

______________ 31. Jumps over a hockey stick while walking around the room.

______________ 32. Puts a hockey stick in a wagon and pulls it around (looking at striking with a stick).

______________ 33. Picks up a ball and looks around the room.

______________ 34. Child swings on the tube.

______________ 35. Child pushes someone else on the tube.

______________ 36. No hockey skills being taught, hits someone with a hockey stick.

______________ 37. Leans up against the side of the orange incline.

______________ 38. Gives an initial push to someone on a scooter at top of the incline.

______________ 39. One foot on scooter, pushing it around the room (not being taught any scooter skills).

______________ 40. One foot on scooter, pushing it around the room.

APPENDIX F

RESULTS OF ANALYSES OF INDIVIDUAL CATEGORY SENSITIVITY TO CHANGE

Category	S_1	S_2	S_3	S_4	S_5	S_6	S_7
Touch(A Skills)	no	no	no	yes	no	no	yes
(B Skills)	no	no	no	no	no	no	no
Attend	no	yes	yes	no	yes	yes	yes
	no	no	yes	no	no	yes	yes
Physical Contact	no	yes	no	yes	no	no	yes
	no	no	no	yes	no	no	yes
Object Manipulation	yes	yes	yes	yes	no	no	yes
	no	no	no	yes	no	no	yes
Below Last Level of Instruction	yes	yes	yes	yes	no	no	yes
	no	no	no	no	no	yes	no
At Last Level of Instruction	no	no	no	yes	yes	yes	no
	no	no	yes	yes	yes	yes	no
Above Last Level of Instruction	no	no	no	no	no	no	no
	no	no	no	no	no	no	no
Skilled Performance Score	yes	yes	yes	yes	yes	yes	yes
	no	no	yes	yes	yes	yes	no
Composite Score	yes	yes	yes	yes	yes	yes	yes
	no	no	no	yes	no	yes	yes

APPENDIX G

RAW DATA

DAILY PERCENTAGE SCORES FOR EACH SUBJECT

Subject 1

	Treatment A Skills			Treatment B Skills		
Ses-sions	Inverted Hang	Jump Over	Swing on a Swing	Rorward Roll	Roll A-round Bar	Throw
1	0	0	5.4	0	0	.2
2	.7	0	0	0	0	0
3	0	0	0	0	0	0
Mean Base-line	.23	0	1.8			
4	0	0	11.1	0	0	1.4
5	.9	0	0	0	0	0
6	–	–	–	–	–	–
7	.2	0	2.1	0	1.4	.4
8	1.9	0	0	0	.7	0
9	5.0	.3	0	0	5.3	.6
10	.8	.6	0	0	.6	.4
11	2.1	.6	6.1	0	.2	1.5
Mean Base-line				.02	.82	.45
12	2.4	.3	0	0	0	2.1
13	0	0	0	0	0	1.3
14	4.3	0	0	.3	0	.2
15	–	–	–	–	–	–
16	–	–	–	–	–	–
17	–	–	–	–	–	–

Subject 2

	Treatment A Skills			Treatment B Skills		
Sessions	Roll A-Round Bar	Slide	Jump Over	Inverted Hang	Scooter on Incline	Forward Roll
1	0	0	0	0	0	0
2	0	.9	0	0	0	0
3	-	-	-	-	-	-
Mean Baseline	0	.45	0			
4	.8	1.4	0	0	0	0
5	0	.6	.4	0	0	0
6	0	1.4	0	0	0	0
7	0	6.5	0	0	0	0
8	0	.7	.2	0	0	0
9	0	3.8	0	0	0	0
10	0	7.1	0	0	0	.3
11	0	5.2	.2	0	0	0
Mean Baseline				0	0	.03
12	0	1.3	0	0	0	0
13	1.0	1.0	0	.2	1.3	0
14	.4	.7	0	0	0	0
15	0	1.1	0	0	0	0
16	0	2.4	0	0	0	0
17	0	.5	0	0	0	0

Subject 3

	Treatment A Skills			Treatment B Skills		
Sessions	Jump on Tramp.	Jump Over	Slide	Swing on Bar	Throw	Scooter on Floor
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	.5	0	0	0
Mean Baseline	0	0	.17			
4	.6	0	2.1	0	0	3.3
5	1.8	0	2.7	0	0	0
6	0	0	1.1	0	0	0
7	0	.8	2.2	0	.3	0
8	0	0	0	0	.2	0
9	.4	0	.4	2.4	0	0
10	-	-	-	-	-	-
11	0	0	1.1	0	0	0
Mean Baseline				.24	.05	.33
12	0	.2	1.6	0	1.1	25.6
13	-	-	-	-	-	-
14	0	0	.2	0	0	24.2
15	0	0	1.3	0	..2	12.5
16	0	.3	0	0	.1	49.4
17	0	0	.6	0	0	20.4

Subject 4

	Treatment A Skills			Treatment B Skills		
Sessions	Ascend Ladder	Jump on Tramp.	Swing on Bar	Jump down	Kick	Scooter on Floor
1	.9	0	0	0	.2	0
2	0	0	0	.2	0	0
3	1.6	0	0	.2	.2	0
Mean Baseline	.83	0	0			
4	1.8	1.8	0	0	.6	0
5	0	0	0	.9	0	0
6	0	.6	0	.3	.3	0
7	2.9	.7	.4	.5	0	0
8	0	.5	.8	.8	0	0
9	1.0	1.2	.2	.4	0	0
10	.4	0	.6	.7	0	0
11	1.2	7.6	0	.3	0	0
Mean Baseline				.39	.12	0
12	.5	2.5	0	0	0	2.6
13	.3	0	0	1.3	0	.8
14	1.4	1.2	0	.4	.2	0
15	–	–	–	–	–	–
16	–	–	–	–	–	–
17	–	–	–	–	–	–

Subject 5

Sessions	Treatment A Skills: Ascend Ladder	Jump Over	Jump on Tramp.	Treatment B Skills: Throw	Jump Down	Scooter on Floor
1	0	.2	3.7	.2	.7	0
2	–	–	–	–	–	–
3	0	0	0	.9	.8	.2
Mean Baseline	0	.10	1.85			
4	0	0	.3	0	1.9	0
5	–	–	–	–	–	–
6	0	0	0	0	1.0	0
7	1.1	2.9	0	0	0	0
8	.2	.4	5.6	.2	.2	0
9	–	–	–	–	–	–
10	–	–	–	–	–	–
11	1.7	.6	0	0	.6	0
Mean Baseline				.19	.74	.03
12	1.5	.1	3.3	.3	.2	2.6
13	1.1	.4	0	.2	.3	.5
14	3.5	0	.4	.2	1.3	1.5
15	1.0	.2	0	0	.6	.6
16	0	0	0	0	0	0
17	.9	0	0	.7	.6	0

Subject 6

	Treatment A Skills			Treatment B Skills		
Sessions	Swing on Bar	Slide	Jump Over	Ascend Ladder	Kick	Scooter on Floor
1	-	-	-	-	-	-
2	0	.4	0	0	.4	0
3	.1	4.0	0	0	0	0
Mean Baseline	.05	2.2	0			
4	.2	1.5	.5	0	0	0
5	0	2.0	.2	.7	0	0
6	.2	.7	0	.5	0	0
7	-	-	-	-	-	-
8	.7	4.4	.4	0	0	0
9	.4	1.8	1.2	.2	0	0
10	1.6	.6	.8	0	0	0
11	1.5	2.0	.6	.3	0	0
Mean Baseline				.41	.19	0
12	0	3.2	0	2.5	0	0
13	0	4.7	0	7.8	.2	0
14	1.1	5.3	.8	4.1	0	0
15	-	-	-	-	-	-
16	0	.9	0	0	0	0
17	.2	8.0	0	3.8	0	0

Subject 7

	Treatment A Skills			Treatment B Skills		
Ses-sions	Ascend Ladder	Walk Up Incline	Throw	Swing on Bar	Climb Onto Box	Scooter Floor
1	-	-	-	-	-	-
2	0	0	0	.5	.2	12.6
3	2.2	0	0	0	0	0
Mean Base-line	1.1	0	0			
4	2.5	0	1.9	0	.4	0
5	0	2.2	.6	0	0	0
6	0	0	0	0	.5	0
7	2.4	0	.7	0	0	0
8	.6	2.1	.2	0	0	0
9	-	-	-	-	-	-
10	0	2.0	0	.6	.2	0
11	0	6.3	.3	.6	0	0
Mean Base-line				.19	.14	1.4
12	0	1.5	.6	0	.5	0
13	1.7	.2	.6	0	0	0
14	2.5	2.7	.8	0	0	0
15	0	0	1.5	0	.4	0
16	0	1.5	0	0	0	.3
17	2.0	2.9	.2	0	.2	1.1